



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Πρόγραμμα Μεταπτυχιακών Σπουδών

«ΔΙΚΑΙΟ ΚΑΙ ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ»

Ακαδημαϊκό έτος 2023-2024

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της Ζωής Ντουράκη (Α.Μ.: ΜΔΙ 2231)

**ΟΙ ΑΡΧΕΣ/ΑΠΑΙΤΗΣΕΙΣ ΤΗΣ ΔΙΚΑΙΟΣΥΝΗΣ ΚΑΙ ΤΗΣ
ΙΣΟΤΗΤΑΣ/ΜΗ ΔΙΑΚΡΙΣΗΣ ΚΑΤΑ ΤΗ ΧΡΗΣΗ ΤΕΧΝΗΤΗΣ
ΝΟΗΜΟΣΥΝΗΣ ΣΤΟ ΠΛΑΙΣΙΟ ΤΗΣ ΕΡΓΑΣΙΑΣ/ΤΩΝ ΕΡΓΑΣΙΑΚΩΝ
ΣΧΕΣΕΩΝ**

**THE PRINCIPLES/REQUIREMENTS OF FAIRNESS AND
EQUALITY/NON-DISCRIMINATION WHEN USING ARTIFICIAL
INTELLIGENCE IN THE CONTEXT OF WORK/EMPLOYMENT
RELATIONS**

Επιβλέπουσα:

Καθηγήτρια Λίλιαν Μήτρου

Πειραιάς, Νοέμβριος 2024

Στους γονείς μου

Ευχαριστίες

Ευχαριστώ την καθηγήτριά μου Λίλιαν Μήτρου για την καθοδήγηση και υποστήριξη

Πίνακας περιεχομένων

ΠΕΡΙΛΗΨΗ.....	7
SUMMARY	9
ΕΙΣΑΓΩΓΗ.....	10
ΚΕΦΑΛΑΙΟ 1 – Η ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ.....	17
1.1. Τι είναι η τεχνητή νοημοσύνη.....	17
1.2. Σε ποια είδη διακρίνεται η τεχνητή νοημοσύνη	19
1.3. Ο τρόπος λειτουργίας της τεχνητής νοημοσύνης	22
ΚΕΦΑΛΑΙΟ 2 – ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΑΛΓΟΡΙΘΜΙΚΗΣ ΜΕΡΟΛΗΨΙΑΣ	25
2.1. Η αλγοριθμική μεροληψία.....	25
2.2. Οι πηγές του προβλήματος της αλγοριθμικής μεροληψίας.....	29
2.3. Το πρόβλημα της έμμεσης μεροληψίας (Proxy Discrimination).....	34
ΚΕΦΑΛΑΙΟ 3 – Η ΔΙΚΑΙΟΣΥΝΗ ΚΑΙ ΙΣΟΤΗΤΑ – ΜΗ ΔΙΑΚΡΙΣΗ ΣΤΗΝ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ.....	36
3.1. Η ηθική αρχή της δικαιοσύνης (fairness) στην τεχνητή νοημοσύνη	36
3.2. Η ενσωμάτωση της δικαιοσύνης ήδη από τον σχεδιασμό της τεχνητής νοημοσύνης (Ethics by Design).....	38
3.3. Η ισότητα - μη διάκριση στο δίκαιο	41
3.4. Η Αλγοριθμική Δικαιοσύνη (Algorithmic Fairness)	44
ΚΕΦΑΛΑΙΟ 4– ΕΦΑΡΜΟΓΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΟ ΧΩΡΟ ΤΗΣ ΕΡΓΑΣΙΑΣ ΚΑΙ ΑΛΓΟΡΙΘΜΙΚΗ ΜΕΡΟΛΗΨΙΑ.....	46
4.1. Η Αλγοριθμική διοίκηση της εργασίας.....	46
4.2. Η εργασία μέσω πλατφόρμας.....	50
4.3. Αναζήτηση και πρόσληψη εργαζομένων με τη βοήθεια της τεχνητής νοημοσύνης.....	54
4.4. Πρακτικό παράδειγμα εφαρμογής τεχνητής νοημοσύνης στις εργασιακές σχέσεις	59
ΚΕΦΑΛΑΙΟ 5 – ΤΟ ΝΟΜΙΚΟ ΟΠΛΟΣΤΑΣΙΟ ΓΙΑ ΤΗΝ ΕΞΑΣΦΑΛΙΣΗ ΤΗΣ ΔΙΚΑΙΟΣΥΝΗΣ ΚΑΤΑ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΙΣ ΕΡΓΑΣΙΑΚΕΣ ΣΧΕΣΕΙΣ	67
5.1. Η εφαρμογή της νομοθεσίας απαγόρευσης των διακρίσεων στη μηχανική μάθηση.....	67
5.2. Ο Γενικός Κανονισμός Προστασίας Δεδομένων.....	71

5.3. Ο Κανονισμός για την τεχνητή νοημοσύνη.....	80
5.4. Ο Ν. 4961/2022.....	89
5.5. Η Οδηγία για τη βελτίωση των συνθηκών εργασίας των εργαζομένων σε πλατφόρμες	95
ΣΥΜΠΕΡΑΣΜΑ	100
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	103

ΠΕΡΙΛΗΨΗ

Σήμερα διανύουμε την 4η βιομηχανική επανάσταση, η οποία περιλαμβάνει καινοτόμες και ανατρεπτικές τεχνολογίες. Στο επίκεντρο του ενδιαφέροντος βρίσκεται το τελευταίο διάστημα η τεχνητή νοημοσύνη, μια τεχνολογία η οποία είναι ικανή να αναπαράγει τις γνωστικές λειτουργίες του ανθρώπου. Η τεχνητή νοημοσύνη βρίσκει εφαρμογή και στον ευαίσθητο τομέα της εργασίας, ώστε γίνεται λόγος για αλγοριθμική διοίκηση των εργασιακών σχέσεων. Οι αλγοριθμικές αποφάσεις και συστάσεις παρουσιάζουν πολλαπλά οφέλη, όμως εγείρουν και αρκετούς ηθικούς προβληματισμούς οι οποίοι έχουν να κάνουν μεταξύ άλλων με την αλγοριθμική μεροληψία. Οι αλγόριθμοι που ενσωματώνουν υποσυνείδητες προκαταλήψεις ή βασίζονται σε δεδομένα μεροληπτικά, τείνουν να ευνοούν ορισμένες κοινωνικές ομάδες σε βάρος άλλων, επιτείνοντας έτσι τις ήδη υπάρχουσες αδικίες και ανισότητες στην κοινωνία. Η εξασφάλιση της δικαιοσύνης στους αλγορίθμους αποτελεί ηθική αλλά και νομική υποχρέωση.

Στην παρούσα μελέτη, θα εξετάσουμε την έννοια της τεχνητής νοημοσύνης, τα είδη στα οποία διακρίνεται και τον τρόπο λειτουργίας της. Θα εστιάσουμε ιδιαίτερα στο πρόβλημα της αλγοριθμικής μεροληψίας, αναλύοντας τις αιτίες που οδηγούν σε αυτό, αναφέροντας και την ειδικότερη περίπτωση του φαινομένου της έμμεσης αλγοριθμικής μεροληψίας. Στη συνέχεια, θα μελετήσουμε την έννοια της δικαιοσύνης στα συστήματα τεχνητής νοημοσύνης, όχι μόνο ως νομική απαίτηση αλλά και ως ηθική αρχή και τεχνικό πρότυπο. Μέσα από πρακτικά παραδείγματα, θα παρουσιάσουμε περιπτώσεις όπου η χρήση της τεχνητής νοημοσύνης στις εργασιακές σχέσεις οδήγησε σε φαινόμενα αλγοριθμικής μεροληψίας, αναλύοντας και μια μελέτη περίπτωσης που αναδεικνύει τις προκλήσεις για την επίτευξη δίκαιων αλγοριθμικών αποφάσεων. Επιπλέον, θα εξετάσουμε το υφιστάμενο νομοθετικό πλαίσιο που αποσκοπεί στην προστασία των εργαζομένων από άδικες αλγοριθμικές αποφάσεις. Αυτό περιλαμβάνει τον Γενικό Κανονισμό Προστασίας Δεδομένων, τη νομοθεσία της Ευρωπαϊκής Ένωσης και της Ελλάδας κατά των διακρίσεων, τον

Κανονισμό για την τεχνητή νοημοσύνη, την Οδηγία για τους εργαζόμενους σε πλατφόρμες και τον νόμο 4961/2022. Θα αναλύσουμε πώς αυτό το πλαίσιο επιδιώκει να εξασφαλίσει τη δικαιοσύνη στις αλγοριθμικές αποφάσεις, τόσο με προληπτικά όσο και με κατασταλτικά μέσα. Τέλος, θα παρουσιάσουμε συμπεράσματα και προβληματισμούς που προέκυψαν από τη μελέτη αυτών των θεμάτων.

SUMMARY

We are currently experiencing the Fourth Industrial Revolution, which includes innovative and disruptive technologies. At the center of attention recently is artificial intelligence (AI), a technology capable of replicating human cognitive functions. AI is also being applied in the sensitive area of work, giving rise to discussions about algorithmic management of labor relations. Algorithmic decisions and recommendations offer multiple benefits, but they also raise significant ethical concerns, particularly related to algorithmic bias. Algorithms that embed subconscious prejudices or are based on biased data tend to favor certain social groups over others, thereby worsening existing injustices and inequalities in society. Ensuring fairness in algorithms is both an ethical and a legal obligation.

In this study, we will examine the concept of artificial intelligence, the categories into which it is divided, and how it functions. We will particularly focus on the issue of algorithmic bias, analyzing the causes that lead to it and mentioning the specific case of indirect algorithmic bias. Next, we will explore the concept of fairness in AI systems, not only as a legal requirement but also as an ethical principle and technical standard. Through practical examples, we will present cases where the use of AI in employment relations has led to instances of algorithmic bias, including a case study that highlights the challenges in achieving fair algorithmic decisions. Additionally, we will review the existing legal framework aimed at protecting workers from unfair algorithmic decisions. This includes the General Data Protection Regulation, anti-discrimination legislation from the European Union and Greece, the AI Act, the Directive on platform workers, and Law 4961/2022. We will analyze how this framework seeks to ensure fairness in algorithmic processes through both preventive and corrective measures. Finally, we will present conclusions and reflections that emerged from the examination of these topics.

ΕΙΣΑΓΩΓΗ

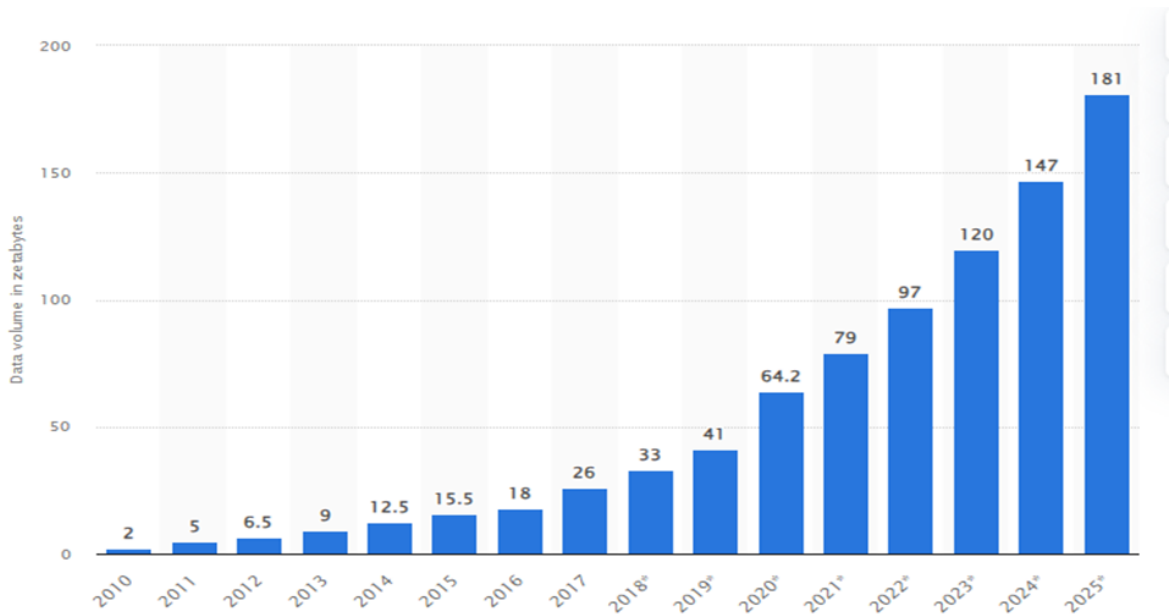
Σήμερα διανύουμε την 4η βιομηχανική επανάσταση, η οποία περιλαμβάνει καινοτόμες και ανατρεπτικές τεχνολογίες, όπως cloud computing, blockchain, IoT και τεχνητή νοημοσύνη. Από τις τεχνολογίες αυτές στο επίκεντρο του ενδιαφέροντος βρίσκεται το τελευταίο διάστημα η τεχνητή νοημοσύνη, μια τεχνολογία η οποία καθιστά τις μηχανές ικανές να «κατανοούν» το περιβάλλον τους, να επιλύουν προβλήματα και να δρουν προς επίτευξη ενός συγκεκριμένου στόχου, αναπαράγοντας ανθρώπινες γνωστικές λειτουργίες, όπως η μάθηση, ο σχεδιασμός και η δημιουργικότητα¹.

Ενώ, λοιπόν, στις προηγούμενες βιομηχανικές επαναστάσεις κινητήριος δύναμη ήταν ο ατμός, το ηλεκτρικό ρεύμα και το πετρέλαιο, τα δεδομένα πλέον θεωρούνται το νέο πετρέλαιο, καθώς αποτελούν την απαραίτητη καύσιμη ύλη για να λειτουργήσει η τεχνητή νοημοσύνη και ένα πολύτιμο αγαθό από το οποίο αντλείται ανυπολόγιστη αξία. Αυτή τη φορά, όμως, η τεχνητή νοημοσύνη, ως νέα μηχανή, δεν έρχεται για να αντικαταστήσει τις σωματικές εργασίες των ανθρώπων αλλά τις νοητικές. Με περίπου 4,66 δισεκατομμύρια ενεργούς χρήστες του Διαδικτύου παγκοσμίως, τα δεδομένα που παράγονται καθημερινά ξεπερνούν κάθε φαντασία. Τα σύνολα των δεδομένων αυτών λόγω του όγκου τους, της ποικιλίας τους και της ταχύτητας με την οποία παράγονται ονομάζονται μεγάλα δεδομένα (Big Data). Σύμφωνα, με έρευνα της Statista, μιας γερμανικής πλατφόρμας δεδομένων και επιχειρηματικής ευφυΐας, έως το 2025 η παγκόσμια παραγωγή δεδομένων προβλέπεται να αυξηθεί δραματικά ξεπερνώντας τα 180 zettabytes². Χαρακτηριστικά 1 zettabyte ισοδυναμεί με ένα

¹Τι είναι η τεχνητή νοημοσύνη και πώς χρησιμοποιείται; (09.09.2020), Ιστοσελίδα Ευρωπαϊκού Κοινοβουλίου [europa.eu](https://www.europarl.europa.eu/news/el/headlines/society/20200827STO85804/ti-einai-i-techniti-noimosuni-kai-pos-chrisimopieitai), <https://www.europarl.europa.eu/news/el/headlines/society/20200827STO85804/ti-einai-i-techniti-noimosuni-kai-pos-chrisimopieitai>, τ.π. 25.01.2024.

² Taylor, P. (16.11.2024), Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2020, with Forecasts from 2021 to 2025, Ιστοσελίδα Statista.com, <https://www.statista.com/statistics/871513/worldwide-data-created/>, τ.π. 25.01.2024.

δισεκατομμύριο terabytes ή με όλους τους κόκκους άμμου όλων των παραλιών στον κόσμο.



Όγκος δεδομένων που δημιουργήθηκαν παγκοσμίως από το 2010 με προβλέψεις για τα έτη 2023 έως 2025 σύμφωνα με τη Statista³.

Ξεπερνώντας τις ανθρώπινες δυνατότητες, η τεχνητή νοημοσύνη έχει την ικανότητα να αναλύει τεράστιους όγκους δεδομένων, να μαθαίνει από αυτά, να εντοπίζει μοτίβα, να κάνει προβλέψεις και να λαμβάνει αποφάσεις. Η αξία που αντλείται από τα δεδομένα είναι ανυπολόγιστη. Η τεχνητή νοημοσύνη είναι μια τεχνολογία gatekeeper, που ως θυρωρός φαίνεται να κατέχει τα κλειδιά για την εκθετική ανάπτυξη της ανθρώπινης ευημερίας⁴. Οι στατιστικές μέθοδοι που ακολουθούνται για να αντληθεί αξία από τα δεδομένα δεν είναι κάτι νέο καθώς ήταν ήδη γνωστές από τα μέσα του 19^{ου} αιώνα. Σήμερα, όμως, το υφιστάμενο θεωρητικό υπόβαθρο, η τεράστια παραγωγή δεδομένων και η υπολογιστική ισχύς καθιστούν δυνατή με τη

³ Taylor, P., ο.π.

⁴ Μήτρου Λ., Μερικές εισαγωγικές σκέψεις, σε συλλογικό έργο Μήτρου Λ., Μπορεί ο αλγόριθμος... να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, Πανεπιστημιακές εκδόσεις Κρήτης, Ηράκλειο Κρήτης, 2023, σελ. 11.

βοήθεια της τεχνητής νοημοσύνης την επεξεργασία των αναρίθμητων αυτών δεδομένων, ώστε να γίνονται προβλέψεις και να επιλύονται προβλήματα προς όφελος της οικονομίας και της ανθρωπότητας.

Έχοντας μπει ενεργά στην ζωή μας με εφαρμογές όπως οι έξυπνοι βοηθοί Siri και Alexa, οι προτάσεις σε υπηρεσίες ροής όπως Netflix και YouTube και τα φίλτρα ανεπιθύμητης ηλεκτρονικής αλληλογραφίας, η τεχνητή νοημοσύνη εξελίσσεται ταχύτατα και αναμένεται να επηρεάσει ριζικά όλους τους τομείς της ανθρώπινης δραστηριότητας (υγεία, μεταφορές, περιβάλλον, εμπόριο, εκπαίδευση, γεωργία, ασφάλεια, δημόσια διοίκηση κλπ) με αποτέλεσμα κράτη και επιχειρήσεις να επενδύουν όλο και περισσότερο στην ανάπτυξη συστημάτων τεχνητής νοημοσύνης. Η ΕΕ προσδοκώντας να αποτελέσει παγκόσμιο κόμβο για αξιόπιστη τεχνητή νοημοσύνη έχει καταρτίσει ευρωπαϊκή στρατηγική για την τεχνητή νοημοσύνη⁵, θεσπίζοντας μάλιστα και τον πρώτο ολοκληρωμένο νόμο παγκοσμίως για τη ρύθμισή της ((ΕΕ) 2024/1689 Κανονισμός για την τεχνητή νοημοσύνη - AI Act)⁶⁷. Στην Ελλάδα η εθνική στρατηγική για τον ψηφιακό μετασχηματισμό της χώρας όπως αποτυπώνεται στη βίβλο ψηφιακού μετασχηματισμού 2020 – 2025 περιλαμβάνει την αξιοποίηση τεχνολογιών τεχνητής νοημοσύνης, ενώ σύντομα η Ελλάδα θα ανήκει στις λίγες ευρωπαϊκές χώρες που διαθέτουν υπερευπολογιστή, ο οποίος θα εξασφαλίζει την απαραίτητη υπολογιστική ισχύ για υλοποίηση έργων τεχνητής νοημοσύνης.

⁵Ευρωπαϊκή προσέγγιση της τεχνητής νοημοσύνης (χ.η.), Ιστοσελίδα της Ευρωπαϊκής Επιτροπής [commission.europa.eu, https://digital-strategy.ec.europa.eu/el/policies/european-approach-artificial-intelligence](https://digital-strategy.ec.europa.eu/el/policies/european-approach-artificial-intelligence), τ.π. 27.03.2024

⁶ Πράξη Τεχνητής Νοημοσύνης της ΕΕ: πρώτος κανονισμός για την τεχνητή νοημοσύνη (09.06.2023), Ιστοσελίδα του Ευρωπαϊκού Κοινοβουλίου [europarl.europa.eu, https://www.europarl.europa.eu/topics/el/article/20230601STO93804/praxi-techniti-noimosunis-tis-ee-protos-kanonismos-gia-tin-techniti-noimosuni](https://www.europarl.europa.eu/topics/el/article/20230601STO93804/praxi-techniti-noimosunis-tis-ee-protos-kanonismos-gia-tin-techniti-noimosuni), τ.π. 27.03.2024

⁷Κανονισμός (ΕΕ) 2024/1689 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 13ης Ιουνίου 2024, για τη θέσπιση εναρμονισμένων κανόνων σχετικά με την τεχνητή νοημοσύνη και την τροποποίηση των κανονισμών (ΕΚ) αριθ. 300/2008, (ΕΕ) αριθ. 167/2013, (ΕΕ) αριθ. 168/2013, (ΕΕ) 2018/858, (ΕΕ) 2018/1139 και (ΕΕ) 2019/2144 και των οδηγιών 2014/90/ΕΕ, (ΕΕ) 2016/797 και (ΕΕ) 2020/1828 (κανονισμός για την τεχνητή νοημοσύνη), https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=OJ%3AL_202401689, τ.π. 19.07.2024.

Οι αλγοριθμικές αποφάσεις και συστάσεις, αδιαμφισβήτητα, παρουσιάζουν πολλαπλά οφέλη, καθώς στηρίζονται σε τάσεις και συμπεριφορές που προκύπτουν από την ανάλυση τεράστιων όγκων διδακτικών παρελθοντικών δεδομένων, μειώνουν το ανθρώπινο λάθος και επιταχύνουν τη διαδικασία λήψης αποφάσεων, εξοικονομώντας χρόνο και πόρους. Οι αποφάσεις αυτές, έχουν σημαντικό αντίκτυπο στους ανθρώπους και ολόκληρη την κοινωνία, ιδίως όταν αφορούν τον τομέα της υγείας, της δικαιοσύνης, της κοινωνικής πρόνοιας, της εργασίας και τον χρηματοπιστωτικό τομέα. Πλέον, ζητήματα όπως η ποινική υποτροπή, η παροχή προνοιακών επιδομάτων, η αξιολόγηση αιτήσεων εργασίας και δανείων, η αύξηση μισθού, η απόλυση, το ύψος των ασφαλίσεων, αφήνονται στην κρίση των αλγορίθμων, οι οποίοι ορίζουν σιωπηλά το πεπρωμένο των ανθρώπων.

Με δεδομένο ότι η τεχνητή νοημοσύνη έχει σημαντικό αντίκτυπο στις ζωές των ανθρώπων, εγείρει αρκετούς ηθικούς προβληματισμούς οι οποίοι σχετίζονται με την προστασία των προσωπικών δεδομένων, τη διαφάνεια και επεξηγησιμότητα των αλγοριθμικών αποφάσεων, τη λογοδοσία και την αλγοριθμική μεροληψία, ώστε τα τελευταία χρόνια βρίσκεται στο επίκεντρο του προβληματισμού στις χώρες όπου επικρατεί κράτος δικαίου.

Η αλγοριθμική μεροληψία δεν αποτελεί υποθετική απειλή αλλά είναι ένα πραγματικό φαινόμενο που επηρεάζει ήδη τους ανθρώπους σήμερα. Οι αλγόριθμοι, αντίθετα με όσα θα αναμέναμε, δεν είναι πάντα ουδέτεροι αλλά κληρονομούν και διακρίνουν, μέσω των συστάσεων και αποφάσεών τους, την ανθρώπινη μεροληψία σε βάρος των ασθενέστερων κοινωνικών ομάδων. Οι διακρίσεις ενσωματώνονται στον κώδικα υπολογιστών και στις τεχνολογίες τεχνητής νοημοσύνης στις οποίες βασιζόμαστε καθημερινά, είτε από επιλογή είτε όχι⁸. Έτσι, εμποδίζεται η πρόσβαση όλων των ανθρώπων σε ίσες ευκαιρίες και δικαιώματα και ενισχύεται η κοινωνική αδικία. Ο φόβος της αλγοριθμικής μεροληψίας είναι βάσιμος και επιβραδύνει την

⁸Noble, S., U., (2018) Algorithms of Oppression: How Search Engines Reinforce Racism, New York University Press.

ανάπτυξη της τεχνητής νοημοσύνης. Όπως χαρακτηριστικά έχει αναφέρει ο John Giannandrea, ο οποίος ηγείτο της τεχνητής νοημοσύνης στην Google, η τεχνητή νοημοσύνη που έχει διαμορφωθεί βασιζόμενη στην ανθρώπινη μεροληψία είναι πολύ πιο επικίνδυνη από την πιθανή δημιουργία ρομπότ που θα σκοτώνουν⁹.

Τα τελευταία χρόνια η τεχνητή νοημοσύνη έχει εισχωρήσει δυναμικά και στον ευαίσθητο χώρο της εργασίας ώστε να γίνεται λόγος για «αλγοριθμική διοίκηση των εργασιακών σχέσεων». Οι εργοδότες με τη βοήθεια της τεχνητής νοημοσύνης αυτοματοποιούν διαδικασίες αλλά λαμβάνουν και καλύτερες αποφάσεις σχετικά με τη λειτουργία της επιχείρησής τους και τη διαχείριση των ανθρωπίνων πόρων. Τα συστήματα λήψης αυτοματοποιημένων αποφάσεων επηρεάζουν σημαντικά τις μελλοντικές προοπτικές σταδιοδρομίας και τα μέσα βιοπορισμού των εργαζομένων αφού πλέον οι αλγόριθμοι ελέγχουν, αξιολογούν, ανταμείβουν και επιπλήττουν τους εργαζόμενους σαν «από μηχανής αφεντικά»¹⁰. Το πρόβλημα της αλγοριθμικής μεροληψίας εντείνεται κατά την αλγοριθμική διοίκηση των εργασιακών σχέσεων, καθώς οι αλγόριθμοι μπορεί να εδραιώσουν ή και να επιτείνουν υφιστάμενες ανισότητες που πλήττουν παραδοσιακά τον χώρο της εργασίας. Η χρήση των αλγορίθμων συχνά είναι αόρατη για τους εργαζόμενους. Η συγκέντρωση κρίσιμων πληροφοριών στη σφαίρα επιρροής των εργοδοτών, η εγγενής αδιαφάνεια στη λειτουργία ορισμένων αλγορίθμων (black box μοντέλα) και η προστασία τους μέσω του εμπορικού απορρήτου καθιστούν ιδιαίτερα δύσκολη για τους εργαζόμενους την προσφυγή ενώπιον δικαστικών ή αρμοδίων διοικητικών εποπτικών αρχών σε περίπτωση που υφίστανται κατά παράβαση του εργατικού δικαίου δυσμενή αδικαιολόγητη διακριτική μεταχείριση. Η πλήρης αυτοματοποίηση, συχνά, αποκλείει τη δυνατότητα των εργαζομένων να μπορούν να επικοινωνούν με άνθρωπο, να λαμβάνουν σχετική αιτιολόγηση και επεξήγηση της απόφασης και να μπορούν να

⁹ Knight, W., (03/10/2017), Forget killer robots—bias is the real AI danger, Ιστοσελίδα MIT Technology Review, technologyreview.com, <https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger/>, τ.π. 27.03.2024.

¹⁰ Aloisi A., Από μηχανής αφεντικό: Οι εργοδοτικές εξουσίες στους χώρους εργασίας που διευθύνονται από αλγόριθμους και τεχνητή νοημοσύνη, Επιθεώρηση Εργατικού Δικαίου, τόμος 82ος, τεύχος 6, έτος 2023.

αμφισβητούν την ορθότητά της, ζητώντας τη διόρθωσή της. Ένα πλέγμα νομοθεσιών, κανόνων δεοντολογίας και μεθοδολογιών της επιστήμης των υπολογιστών έρχεται να προστατεύσει τους εργαζόμενους ώστε να εξασφαλίζεται η διαφάνεια, η δικαιοσύνη, η ανθρώπινη εποπτεία και η λογοδοσία κατά την αλγοριθμική διοίκηση των εργασιακών σχέσεων. Οι προκλήσεις ωστόσο παραμένουν και το φαινόμενο της αλγοριθμικής μεροληψίας απαιτεί επαγρύπνηση και συνεχή παρακολούθηση. Είναι βέβαιο ότι η τεχνητή νοημοσύνη αναμένεται να αποτελέσει μείζον ζήτημα για τα ανθρώπινα δικαιώματα στον εικοστό πρώτο αιώνα, ενώ πρόσφατα αρχίζουμε να κατανοούμε τις μακροπρόθεσμες συνέπειες αυτών των εργαλείων λήψης αποφάσεων τόσο στην συγκάλυψη όσο και στην εμβάθυνση των κοινωνικών ανισοτήτων¹¹.

Στην παρούσα μελέτη, λοιπόν, θα εξετάσουμε την έννοια της τεχνητής νοημοσύνης, τα είδη στα οποία διακρίνεται, και τον τρόπο λειτουργίας της. Θα εστιάσουμε ιδιαίτερα στο πρόβλημα της αλγοριθμικής μεροληψίας, αναλύοντας τις αιτίες που οδηγούν σε αυτό, αναφέροντας και την ειδικότερη περίπτωση του φαινομένου της έμμεσης αλγοριθμικής μεροληψίας. Στη συνέχεια, θα μελετήσουμε την έννοια της δικαιοσύνης στα συστήματα τεχνητής νοημοσύνης, όχι μόνο ως νομική απαίτηση αλλά και ως ηθική αρχή και τεχνικό πρότυπο. Μέσα από πρακτικά παραδείγματα, θα παρουσιάσουμε περιπτώσεις όπου η χρήση τεχνητής νοημοσύνης στις εργασιακές σχέσεις οδήγησε σε φαινόμενα αλγοριθμικής μεροληψίας, αναλύοντας και μια μελέτη περίπτωσης που αναδεικνύει τις προκλήσεις για την επίτευξη δίκαιων αλγοριθμικών αποφάσεων. Επιπλέον, θα εξετάσουμε το υφιστάμενο νομοθετικό πλαίσιο που αποσκοπεί στην προστασία των εργαζομένων από άδικες αλγοριθμικές αποφάσεις. Αυτό περιλαμβάνει τον Γενικό Κανονισμό Προστασίας Δεδομένων, τη νομοθεσία της Ευρωπαϊκής Ένωσης και της Ελλάδας κατά των διακρίσεων, τον Κανονισμό για την τεχνητή νοημοσύνη, την Οδηγία για τους εργαζόμενους σε πλατφόρμες και τον νόμο 4961/2022. Θα αναλύσουμε πώς αυτό το πλαίσιο επιδιώκει να εξασφαλίσει τη δικαιοσύνη στις αλγοριθμικές αποφάσεις, τόσο με προληπτικά όσο

¹¹ Noble, S. U., ο.π.

και με κατασταλτικά μέσα. Τέλος, θα παρουσιάσουμε συμπεράσματα και προβληματισμούς που προέκυψαν από τη μελέτη αυτών των θεμάτων.

ΚΕΦΑΛΑΙΟ 1 – Η ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

Η κατανόηση του τι είναι η τεχνητή νοημοσύνη και πως λειτουργεί είναι σημαντική προκειμένου να αντιληφθούμε πως προκύπτει το πρόβλημα των διακρίσεων και της μεροληψίας στις αλγοριθμικές αποφάσεις και συστάσεις αλλά και πώς αυτό μπορεί να εξαλειφθεί, τηρουμένης της αρχής της δικαιοσύνης.

1.1. Τι είναι η τεχνητή νοημοσύνη

Η απαρχή της τεχνητής νοημοσύνης τοποθετείται ιστορικά στον Οκτώβρη του 1950, όταν ο Alan Turing Βρετανός μαθηματικός, κρυπτογράφος και πατέρας της τεχνητής νοημοσύνης, εκδίδει το άρθρο του *Υπολογιστικά Μηχανήματα και Νοημοσύνη (Computing Machinery and Intelligence)*¹², το οποίο ξεκινούσε με την ερώτηση «Μπορούν οι μηχανές άραγε να σκέπτονται;». Προκειμένου να απαντήσει το ερώτημα πρότεινε το «παιχνίδι της μίμησης», γνωστό πλέον ως «δοκιμασία Turing» (Turing Test). Στο παιχνίδι αυτό ένας παίκτης σε ρόλο «ανακριτή», ενώ είναι απομονωμένος από τους άλλους δύο παίκτες, έναν άνθρωπο και έναν υπολογιστή, προσπαθεί μέσω γραπτής επικοινωνίας και με ερωτήσεις να διακρίνει τον υπολογιστή από τον άνθρωπο. Αν ο «ανακριτής» δυσκολευτεί να διακρίνει τον υπολογιστή, τότε ίσως έχουμε να κάνουμε με έναν υπολογιστή που σκέφτεται. Έτσι για να κριθεί αν ένας υπολογιστής είναι ευφυής, θα πρέπει να περάσει το τεστ και να μιμηθεί με επιτυχία τον άνθρωπο. Ιστορικά ο όρος τεχνητή νοημοσύνη (AI-Artificial Intelligence) χρησιμοποιήθηκε για πρώτη φορά το 1955 από μια ομάδα που αποτελούνταν από πέντε επιστήμονες της πληροφορικής τους John McCarthy, Marvin Minsky, Nathaniel Rochester και Claude Shannon για να περιγράψει «την επιστήμη και μηχανική της κατασκευής ευφυών μηχανών»¹³.

¹²Turing, A.M. (1950) Computing Machinery and Intelligence. Mind, 59, 433-460, <https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf>, τ.π. 30.01.2023.

¹³ Dutceac Segesten A., The definition of AI, course AI, Business & the Future of Work Lund University, Coursera.org, <https://www.coursera.org/learn/ai-business-future-of-work/lecture/MAOQw/the-definition-of-ai>, τ.π. 27.03.2024

Σήμερα οι περισσότεροι άνθρωποι αντιλαμβάνονται την τεχνητή νοημοσύνη έτσι όπως παρουσιάζεται στις ταινίες, δηλαδή ως ανθρωπόμορφα ρομπότ που σκέφτονται και μαθαίνουν σαν άνθρωποι. Το γεγονός αυτό μπορεί να έχει αρνητικές συνέπειες όταν εκτιμώνται τα οφέλη και οι κίνδυνοι της τεχνητής νοημοσύνης, δημιουργώντας υψηλές προσδοκίες για μια τεχνολογία που στην πραγματικότητα, προς το παρόν, έχει πιο περιορισμένες δυνατότητες, εστιάζοντας τους κινδύνους της τεχνητής νοημοσύνης στα ρομπότ που οπλοφορούν και στους υπερφυείς πράκτορες και υποτιμώντας άλλα πιο ρεαλιστικά προβλήματα όπως η αλγοριθμική μεροληψία και η έλλειψη διαφάνειας¹⁴.

Η τεχνητή νοημοσύνη δεν είναι μια σαφώς ορισμένη τεχνολογία, ενώ κάθε κατηγοριοποίηση της είναι εξ ορισμού προσωρινή καθώς πιθανότατα να μην αποδίδει επαρκώς την τεχνολογική πραγματικότητα σε δέκα χρόνια από σήμερα. Το κοινό σημείο των διαφόρων ορισμών και προσεγγίσεων εστιάζει στη συσχέτιση των συστημάτων τεχνητής νοημοσύνης με ανθρώπινες ή έλλογες ιδιότητες¹⁵.

Οι ορισμοί που αποδίδονται σήμερα στην τεχνητή νοημοσύνη προέρχονται τόσο από επιστήμονες της πληροφορικής όσο και από τον νομοθέτη. Σύμφωνα, λοιπόν, με τη Microsoft η τεχνητή νοημοσύνη είναι μια ομάδα τεχνολογιών που επιτρέπουν στους υπολογιστές να αντιλαμβάνονται, να μαθαίνουν, να επιχειρηματολογούν και να συνδράμουν στη λήψη αποφάσεων για την επίλυση προβλημάτων με τρόπους που μοιάζουν με αυτά που κάνουν οι άνθρωποι¹⁶. Ενώ, σύμφωνα με τον ορισμό που πρότεινε ο ΟΟΣΑ¹⁷ και υιοθετήθηκε από τον Κανονισμό για την τεχνητή νοημοσύνη, ένα σύστημα τεχνητής νοημοσύνης είναι ένα μηχανικό σύστημα που έχει σχεδιαστεί

¹⁴Dutceac Segesten, A., ο.π.

¹⁵ Μήτρου Λ., Μερικές εισαγωγικές σκέψεις, σε συλλογικό έργο Μήτρου Λ. Μπορεί ο αλγόριθμος... να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, Πανεπιστημιακές εκδόσεις Κρήτης, Ηράκλειο Κρήτης, 2023, σελ. 14- 15.

¹⁶ Κουσοσύνη – Πανταζοπούλου Α., Νομικές διαστάσεις της τεχνητής νοημοσύνης (παρόν και μέλλον), ΕλλΔνη 1/2019, σελ. 313.

¹⁷ Σύμφωνα με τον ορισμό του ΟΟΣΑ «An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments», βλ. ιστοσελίδα ΟΟΣΑ <https://oecd.ai/en/wonk/definition> τ.π. 04.08.2024.

για να λειτουργεί με διαφορετικά επίπεδα αυτονομίας και μπορεί να παρουσιάζει προσαρμοστικότητα μετά την εφαρμογή του και το οποίο για ρητούς ή σιωπηρούς στόχους, συνάγει, από τα στοιχεία εισόδου που λαμβάνει, πώς να παράγει στοιχεία εξόδου, όπως προβλέψεις, περιεχόμενο, συστάσεις ή αποφάσεις, που μπορούν να επηρεάσουν υλικά ή εικονικά περιβάλλοντα¹⁸. Ο τελευταίος ορισμός είναι επιτυχής καθώς είναι τεχνολογικά ακριβής, σε κατανοητή γλώσσα, παρέχει παραδείγματα λειτουργιών και εστιάζει στις υπάρχουσες δυνατότητες της τεχνητής νοημοσύνης¹⁹.

Σε κάθε περίπτωση η επίτευξη ενός κοινά αποδεκτού ορισμού ο οποίος θα είναι ακριβής και κατανοητός είναι πολύ σημαντική ώστε να υπάρχει καλύτερη συνεργασία μεταξύ των επιστημόνων της πληροφορικής, των νομοθετών και των υπευθύνων χάραξης πολιτικής αλλά και να διευκολύνονται οι φορείς κατά την εφαρμογή των κανονισμών και νομοθεσιών που σχετίζονται με την τεχνητή νοημοσύνη²⁰.

1.2. Σε ποια είδη διακρίνεται η τεχνητή νοημοσύνη

Η τεχνητή νοημοσύνη διακρίνεται σε τρία είδη: α) την αδύναμη τεχνητή νοημοσύνη (Weak AI/ Specialized AI/ Narrow AI), β) την ισχυρή τεχνητή νοημοσύνη (Strong AI /Artificial General Intelligence) και γ) την τεχνητή υπερνοημοσύνη (Artificial superintelligence).

Όλες οι εφαρμογές τεχνητής νοημοσύνης που υπάρχουν σήμερα θεωρείται ότι εντάσσονται στην αδύναμη τεχνητή νοημοσύνη, η οποία παρά την ονομασία της, κάθε άλλο παρά αδύναμη είναι. Η αδύναμη τεχνητή νοημοσύνη εστιάζει σε μια συγκεκριμένη εργασία προκειμένου να την τελειοποιήσει και να την εκτελέσει καλύτερα από τους ανθρώπους. Η αδύναμη τεχνητή νοημοσύνη μπορεί να λύσει

¹⁸ Art. 3 AI Act: AI system means a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

¹⁹ Dutceac Segesten A., ο.π.

²⁰ Dutceac Segesten, A., ο.π.

δύσκολα προβλήματα πολύ γρήγορα, αλλά διαθέτει συγκεκριμένο επίπεδο «νοημοσύνης» στο πλαίσιο ενός αυστηρά καθορισμένου τομέα, χωρίς να διαθέτει «αντίληψη» για οτιδήποτε άλλο εκτός από τα δεδομένα που της παρέχουν οι δημιουργοί της και στο πλαίσιο της καθορισμένης από αυτούς λειτουργίας²¹. Ένα από τα πρώτα, πιο διάσημα παραδείγματα αδύναμης τεχνητής νοημοσύνης είναι ο υπολογιστής της IBM «Deep Blue», ο οποίος κατάφερε να κερδίσει τον πρωταθλητή σκακιού Gary Kasparov σε σκακιστικό αγώνα το 1997, προβλέποντας 20 κινήσεις μπροστά από τον αντίπαλό του, κατόρθωμα που κανένας άνθρωπος μέχρι σήμερα δεν έχει καταφέρει. Άλλα παραδείγματα αδύναμης τεχνητής νοημοσύνης είναι τα Chatbots, τα φίλτρα ανεπιθύμητης ηλεκτρονικής αλληλογραφίας, οι έξυπνοι βοηθοί όπως οι Siri, Alexa και Cortana, τα αυτόνομα αυτοκίνητα και οι χάρτες της Google. Στην ίδια κατηγορία εντάσσονται και οι εφαρμογές της αλγοριθμικής διοίκησης της εργασίας, που θα αναφέρουμε αργότερα στην παρούσα, όπως για παράδειγμα μια εφαρμογή που βοηθά τη διαδικασία των προσλήψεων και μπορεί να προβλέψει εάν ένας υποψήφιος εργαζόμενος αναμένεται να έχει καλή απόδοση σε μια θέση εργασίας αν προσληφθεί.

Η ισχυρή τεχνητή νοημοσύνη ή αλλιώς γενική τεχνητή νοημοσύνη είναι η τεχνητή νοημοσύνη που θα είναι ικανή να εκτελεί ενέργειες με τον ίδιο τρόπο όπως ο άνθρωπος, μιμούμενη την ανθρώπινη γενική νοημοσύνη, λύνοντας προβλήματα και αποκτώντας δεξιότητες. Η ισχυρή τεχνητή νοημοσύνη θα μπορεί να γενικεύσει τη γνώση και να την εφαρμόσει από τη μια εργασία στην άλλη, να προγραμματίσει εκ των προτέρων σύμφωνα με τις τρέχουσες γνώσεις και να προσαρμοστεί σε ένα περιβάλλον το οποίο μεταβάλλεται. Αυτού του είδους η τεχνητή νοημοσύνη θα μπορεί να συλλέγει αυτόνομα ερεθίσματα και δεδομένα από το περιβάλλον, να ενεργεί αυτόνομα αλλά και να εκπαιδεύεται από αυτά τα δεδομένα. Παραδείγματα ισχυρής τεχνητής νοημοσύνης συναντάμε μόνο σε ταινίες επιστημονικής φαντασίας

²¹ Μήτρου Λ., Μερικές εισαγωγικές σκέψεις, σε συλλογικό έργο Μήτρου Λ. Μπορεί ο αλγόριθμος... να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, Πανεπιστημιακές εκδόσεις Κρήτης, Ηράκλειο Κρήτης, 2023, σελ. 12 βλ. υποσημείωση 12.

όπως το Her ή Ex Machina, όπου ο πρωταγωνιστής ερωτεύεται ένα πρόγραμμα τεχνητής νοημοσύνης. Ένας υπολογιστής με γενική τεχνητή νοημοσύνη θα μπορούσε να σαρώσει όλη τη γνώση του κόσμου που υπάρχει στο Διαδίκτυο για να λύσει μερικά από τα πιο δύσκολα προβλήματα ή ακόμα και να τα προβλέψει και να τα αντιμετωπίσει προτού δημιουργηθούν²². Η ισχυρή τεχνητή νοημοσύνη ίσως και να μην υλοποιηθεί ποτέ. Τα τελευταία χρόνια έχουν δημιουργηθεί τα λεγόμενα μεγάλα γλωσσικά μοντέλα (Large Language Models ή Language Learning Models - LLM) όπως το Chat GPT της Open AI και το Bert της Google τα οποία είναι ένα είδος αλγορίθμου τεχνητής νοημοσύνης που χρησιμοποιεί τεχνικές βαθιάς μάθησης και μαθαίνει από τεράστιους όγκους δεδομένων για να παράγει νέο περιεχόμενο όπως για παράδειγμα μια συστατική επιστολή, ένα δικόγραφο ή μια εικόνα. Αυτή η κατηγορία τεχνητής νοημοσύνης ονομάζεται γενετική – δημιουργική τεχνητή νοημοσύνη (Generative AI), ανήκει στη αδύναμη τεχνητή νοημοσύνη και προσεγγίζει κάπως θα λέγαμε την ισχυρή τεχνητή νοημοσύνη αφού μοιάζει να κατανοεί όπως ένας άνθρωπος τις ερωτήσεις που δέχεται και να παρέχει κατάλληλες απαντήσεις. Βέβαια, ή αλήθεια είναι ότι στην πραγματικότητα η γενετική – δημιουργική τεχνητή νοημοσύνη λειτουργεί βάσει των μοτίβων που εντοπίζει χωρίς να μπορεί να κατανοήσει τις πληροφορίες που δέχεται, δεν διαθέτει «ανθρώπινη» αντίληψη, ούτε μπορεί αυτόνομα να εκτελέσει νέα καθήκοντα χωρίς εκ νέου εκπαίδευση.

²² Glover, E. (29.09.2022), Strong AI vs. Weak AI: What's the difference?, builtin.com, <https://builtin.com/artificial-intelligence/strong-ai-weak-ai>, τ.π. 27.03.2024



Εικόνα που δημιουργήθηκε από το DALL-E εργαλείο Generative AI για τη σύνθεση εικόνων

Τέλος, η υπερνοημοσύνη (Artificial Superintelligence) αποτελεί το τρίτο είδος της τεχνητής νοημοσύνης, θα αποτελέσει μετεξέλιξη της γενικής τεχνητής νοημοσύνης και θα διαθέτει γνωστικές ικανότητες που θα ξεπερνούν αυτές των πιο έξυπνων ανθρώπων σε όλον τον κόσμο.

1.3. Ο τρόπος λειτουργίας της τεχνητής νοημοσύνης

Όπως αναφέρει ο Thomas Davenport στο άρθρο του «From analytics to artificial intelligence», οι περισσότερες εφαρμογές τεχνητής νοημοσύνης βασίζονται σε αλγόριθμους και αναλυτικές μαθηματικές και στατιστικές μεθόδους²³. Ένας αλγόριθμος είναι μια συνταγή ή διαφορετικά μια σειρά από εντολές που αποτελείται από λογική και μαθηματικές και στατιστικές μεθόδους, οι οποίες αποσκοπούν να λύσουν ένα πρόβλημα, όπως για παράδειγμα ποιοι εργαζόμενοι αν προσληφθούν θα είναι αποδοτικοί. Ο αλγόριθμος λειτουργεί σαν ένα σύστημα που τροφοδοτείται από δεδομένα, τα αναλύει και εξάγει ένα αποτέλεσμα. Οι αλγόριθμοι χρησιμοποιούνται

²³Davenport, T. H., From analytics to artificial intelligence, Journal of Business Analytics Volume 1, 2018, 19/11/2018,tandonline.com, <https://www.tandfonline.com/doi/full/10.1080/2573234X.2018.1543535>, τ.π. 27.03.2024

συχνά για να κάνουν προβλέψεις. Για κάθε συγκεκριμένη εργασία, ένας αλγόριθμος τροφοδοτείται με δεδομένα με τα οποία εκπαιδεύεται και έτσι δημιουργεί ένα μοντέλο που χρησιμοποιείται στην πράξη για μια εργασία στον πραγματικό κόσμο²⁴.

Η πιο γνωστή κατηγορία αλγορίθμων που σχετίζεται με την τεχνητή νοημοσύνη είναι αυτή της μηχανικής μάθησης. Η μηχανική μάθηση είναι η διαδικασία κατά την οποία οι υπολογιστές εκπαιδεύονται χρησιμοποιώντας μεγάλες ποσότητες δεδομένων «εκπαίδευσης», όπως κείμενα, εικόνες και ήχο, ώστε να μάθουν να ολοκληρώνουν ανεξάρτητα εργασίες, οι οποίες συνδέονται με την ανθρώπινη ευφυΐα, όπως μια σύσταση. Στη μηχανική μάθηση συναντάμε δυο τεχνικές, την εποπτευόμενη μάθηση ή αλλιώς προβλεπτική αναλυτική, η οποία κάνει προβλέψεις για το μέλλον και τη μη εποπτευόμενη μάθηση ή αλλιώς ανακάλυψη προτύπων, η οποία εξηγεί γεγονότα του παρελθόντος. Η πρώτη προϋποθέτει την ύπαρξη ενός ανθρώπου ως επόπτη δασκάλου ο οποίος τροφοδοτεί τον υπολογιστή με χαρακτηρισμένα δεδομένα - labeled data ή παραδείγματα προβλημάτων και λύσεων. Έτσι για να εκπαιδευτεί ένας υπολογιστής να αναγνωρίζει την εικόνα μιας γάτας και ενός σκύλου ο άνθρωπος επόπτης θα πρέπει να τροφοδοτήσει τον υπολογιστή με πολλές εικόνες που έχει χαρακτηρίσει ο ίδιος ως εικόνες «γάτας» ή «σκύλου» (μεταβλητή στόχος), δηλαδή παραδείγματα με γάτες και σκύλους. Ο υπολογιστής θα επεξεργαστεί τα χαρακτηριστικά αυτών των εικόνων, δηλαδή τα pixel αυτών, εντοπίζοντας στις εικόνες μοτίβα, κανόνες, τάσεις ή κάποιο χαρακτηριστικό που εμφανίζεται ξανά και ξανά με έναν προβλέψιμο τρόπο (όπως σχήματα, χρώματα, υφές) ώστε να μπορεί μετά την εκπαίδευσή του αυτή να αναγνωρίζει νέες εικόνες με σκύλους ή γάτες, αντιλαμβανόμενο ποια χαρακτηριστικά προσδιορίζουν και συνθέτουν μια γάτα ή έναν σκύλο. Ομοίως, σε ένα σύστημα τεχνητής νοημοσύνης που υποβοηθά στις προσλήψεις εργαζομένων, ο υπολογιστής θα μελετήσει τα μοτίβα στα ιστορικά δεδομένα παρελθοντικών προσλήψεων για να κατανοήσει ποιες ιδιότητες ή εμπειρίες τείνουν να οδηγήσουν σε επιτυχημένες προσλήψεις υπαλλήλων υψηλών επιδόσεων.

²⁴ Fundamental Rights Agency. (2022). Bias in algorithms, artificial intelligence, and discrimination. Luxembourg Publications Office, <https://fra.europa.eu/en/publication/2022/bias-algorithm> τ.π. 16.11.2024

Το μοντέλο θα χρησιμοποιήσει τα μοτίβα – κανόνες που έμαθε για να φιλτράρει νέες αιτήσεις και να προτείνει υποψήφιους που είναι πιο πιθανό να ταιριάζουν στη θέση. Ένα σύστημα τεχνητής νοημοσύνης δεν ακολουθεί μόνο προκαθορισμένους από τον άνθρωπο κανόνες αλλά βρίσκει νέους κανόνες, οι οποίοι μπορεί να μην ήταν αντιληπτοί από τον αναλυτή δεδομένων. Έτσι, χάρη στα χαρακτηριστικά της μηχανικής μάθησης οι αλγόριθμοι μπορούν με έναν ορισμένο βαθμό αυτονομίας και ελάχιστη ανθρώπινη επίβλεψη να επιλέγουν χρήσιμα αποτελέσματα εντοπίζοντας μοτίβα στα υπάρχοντα δεδομένα προκειμένου να δημιουργήσουν μοντέλα που προβλέπουν μελλοντικά αποτελέσματα²⁵. Η τεχνητή νοημοσύνη διαφέρει από τις κλασικές τεχνολογίες αυτοματισμού κατά το ότι δεν παράγει αποφάσεις επί τη βάση μιας προ-προγραμματισμένης λογικής τύπου «εάν – τότε», όπου οι οδηγίες αναφορικά με το τι εισέρχεται παράγουν ακριβώς το ίδιο αποτέλεσμα. Αντίθετα η τεχνητή νοημοσύνη χαρακτηρίζεται από αυτονομία στη λειτουργία και στην εκμάθηση.

Στην περίπτωση της εποπτευόμενης μάθησης ο επόπτης μπορεί να ελέγχει την απόδοση του μοντέλου διαπιστώνοντας πόσες περιπτώσεις ταξινομήθηκαν σωστά. Αντίθετα στην μη εποπτευόμενη μάθηση δεν δίνονται δεδομένα που έχουν επισημανθεί ή οδηγίες στον υπολογιστή αλλά ο υπολογιστής προσπαθεί να βρει τα μοτίβα που υπάρχουν στα δεδομένα χωρίς καμία ανθρώπινη βοήθεια. Η μη εποπτευόμενη μάθηση επιχειρεί να βρει τις εγγενείς ομοιότητες μεταξύ διαφορετικών περιπτώσεων. Παράδειγμα μη εποπτευόμενης μάθησης αποτελεί η τμηματοποίηση πελατών σε ομάδες με βάση την παρελθοντική καταναλωτική τους συμπεριφορά.

Οι σημαντικότερες τεχνικές - αλγόριθμοι για να λυθεί ένα πρόβλημα προβλεπτικής αναλυτικής είναι τα δέντρα αποφάσεων, η παλινδρόμηση (regression), τα νευρωνικά δίκτυα, οι μηχανές υποστήριξης διανυσμάτων (support vector machines) και τα τυχαία δάση (random forests). Κάποιες από αυτές τις τεχνικές έχουν την δυνατότητα να εξηγούν στους χρήστες τον τρόπο με τον οποίο εξάγουν ένα αποτέλεσμα, όπως είναι

²⁵ Aloisi A., ο.π.

για παράδειγμα τα δέντρα αποφάσεων και η παλινδρόμηση, ενώ κάποιες άλλες όπως τα νευρωνικά δίκτυα, οι μηχανές υποστήριξης διανυσμάτων (support vector machines) και τα τυχαία δάση (random forests) θεωρούνται black box μοντέλα διότι δεν παρέχουν κάποια επεξηγησιμότητα της απόφασης²⁶. Ωστόσο, στο σημείο αυτό θα πρέπει να τονίσουμε ότι σε αντίθεση με την ευρέως διαδεδομένη παρανόηση, πίσω από τους αλγορίθμους υπάρχουν πάντα άνθρωποι οι οποίοι δεν απαλλάσσονται από την ευθύνη σε περίπτωση παράνομων αποτελεσμάτων, παραβίασης της ιδιωτικής ζωής ή δυσμενών διακρίσεων ούτε καν στην περίπτωση των τεχνικών μηχανικής μάθησης²⁷.

ΚΕΦΑΛΑΙΟ 2 – ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΑΛΓΟΡΙΘΜΙΚΗΣ ΜΕΡΟΛΗΨΙΑΣ

2.1. Η αλγοριθμική μεροληψία

Στη μηχανική μάθηση μόλις εκπαιδευτεί, ένα μοντέλο πάνω στα δεδομένα εκπαίδευσης χρησιμοποιείται για να κάνει προβλέψεις που μπορούν να βοηθήσουν στην αυτοματοποίηση των αποφάσεων. Οι αποφάσεις αυτές μπορεί να αφορούν οτιδήποτε όπως μια πρόταση σχετικά με το ποια σειρά πιθανότατα να αρέσει σε έναν θεατή στο Netflix αλλά και άλλα ζητήματα που έχουν σημαντικό αντίκτυπο στις ζωές των ανθρώπων και είναι κρίσιμα για την υγεία, την ελευθερία και την οικονομική ευημερία τους, όπως η διάγνωση ασθενειών, η κρίση περί της ποινικής υποτροπής, η παροχή προνοιακών επιδομάτων, η αξιολόγηση αιτήσεων εργασίας και αιτήσεων δανείων, η αύξηση μισθού, η απόλυση, και το ύψος των ασφαλίσεων.

Η χρήση της τεχνητής νοημοσύνης μπορεί να προσφέρει πολλά οφέλη αλλά μπορεί υπό συγκεκριμένες συνθήκες να πλήξει πανανθρώπινες αξίες και να οδηγήσει μέσω των συστάσεων και αποφάσεων της σε παραβιάσεις θεμελιωδών δικαιωμάτων, όπως αυτό της απαγόρευσης των διακρίσεων λόγω φύλου, φυλετικής ή εθνοτικής

²⁶ Σχετικά με την επιχειρηματική αναλυτική βλ. Klimberg R., Miori, V., (October 2010), Back in Business, https://www.researchgate.net/publication/305958445_Back_in_Business, τ.π. 27.03.2024, Baesens, B., Business analytics: Analytics in a Big Data World (Wiley and SAS Business Series), 1st Edition, Wiley Editions, 2015 και Davenport, T., H., The new world of business analytics, Jan 2010, International Institute for Analytics.

²⁷Aloisi, A., ο.π

καταγωγής, θρησκείας ή πεποιθήσεων, αναπηρίας, ηλικίας ή γενετήσιου προσανατολισμού²⁸.

Ως «μεροληψία στην μηχανική μάθηση» ή «αλγοριθμική μεροληψία» ή «μεροληψία στην τεχνητή νοημοσύνη» νοείται η τάση των αλγορίθμων να αντικατοπτρίζουν τις ανθρώπινες προκαταλήψεις στις προβλέψεις και συστάσεις τους και το φαινόμενο κατά το οποίο οι αλγόριθμοι παράγουν αποτελέσματα που είναι συστημικά προκατειλημμένα σε βάρος υποεκπροσωπούμενων προστατευόμενων κοινωνικών ομάδων, λόγω λανθασμένων υποθέσεων στη διαδικασία μηχανικής μάθησης²⁹ και λανθασμένων – μεροληπτικά επιλεγμένων και μη ποιοτικών δεδομένων (π.χ. λανθασμένα, ελλιπή ή μη αντιπροσωπευτικά δεδομένα εκπαίδευσης καθώς και δεδομένα που ως απότοκα κοινωνικής συμπεριφοράς ενσαρκώνουν προυπάρχουσες πρακτικές διακρίσεων).

Στη μηχανική μάθηση τα δεδομένα εκπαίδευσης είναι κρίσιμα για την απόδοση του μοντέλου ώστε συχνά να αναφέρεται στην επιστήμη των υπολογιστών η φράση “garbage in, garbage out” ή διαφορετικά «ότι δώσεις θα πάρεις». Προκειμένου να γίνει καλύτερα κατανοητό αυτό θα παραθέσουμε ένα παράδειγμα. Έστω λοιπόν, ότι μια τράπεζα αναπτύσσει ένα σύστημα τεχνητής νοημοσύνης για τον υπολογισμό του πιστωτικού κινδύνου των υποψήφιων πελατών της. Η τράπεζα σκοπεύει να χρησιμοποιήσει το σύστημα για να λάβει βοήθεια και υποστήριξη για να εγκρίνει ή να απορρίπτει επικείμενες αιτήσεις δανείων. Το σύστημα εκπαιδεύεται με ένα μεγάλο σύνολο δεδομένων που περιέχει διάφορες πληροφορίες για προηγούμενους δανειολήπτες, όπως το επάγγελμά τους, το εισόδημά τους, την ηλικία τους και το αν αποπλήρωσαν το δάνειό τους. Η τράπεζα υποβάλλει σε δοκιμή το σύστημα και το σύστημα βρίσκεται να αποδίδει χαμηλότερες πιστωτικές βαθμολογίες στις γυναίκες και συνεπώς να μεροληπτεί σε βάρος αυτών λόγω του φύλου τους. Αυτό, λοιπόν,

²⁸Λευκή Βίβλος Τεχνητή Νοημοσύνη - Η ευρωπαϊκή προσέγγιση της αριστείας και της εμπιστοσύνης, Ευρωπαϊκή Επιτροπή 19/02/2020 COM (2020) final

²⁹Gillis A., (χ.η.) Definition of machine learning bias (AI bias), <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-bias-algorithm-bias-or-AI-bias>, τ.π. 28.03.2024.

μπορεί να οφείλεται στην υπερεκροσώπηση των αντρών στο ιστορικό δείγμα δεδομένων εκπαίδευσης μιας και ιστορικά οι γυναίκες απέκτησαν οικονομική ανεξαρτησία και άρα δυνατότητα υποβολής αιτήσεων δανείων πολύ αργότερα από τους άντρες. Το μοντέλο, συνεπώς, λόγω αυτής της υπερεκροσώπησης, αποδίδει περισσότερη προσοχή στις στατιστικές σχέσεις που προβλέπουν ποσοστά αποπληρωμής για τους άνδρες και λιγότερη στις στατιστικές σχέσεις που προβλέπουν τα ποσοστά αποπληρωμής για τις γυναίκες, διότι τις θεωρεί στατιστικά λιγότερο σημαντικές. Επιπλέον, το μεροληπτικό αυτό αποτέλεσμα μπορεί να επέρχεται διότι το σύνολο των δεδομένων εκπαίδευσης αντικατοπτρίζει τις διακρίσεις του παρελθόντος. Έτσι, εάν στο παρελθόν οι αιτήσεις δανείων από γυναίκες απορρίπτονταν πιο συχνά από αυτές των ανδρών λόγω προκαταλήψεων, των υπευθύνων λήψης αποφάσεων δανειοδότησης τότε οποιοδήποτε μοντέλο που βασίζεται σε τέτοια δεδομένα εκπαίδευσης είναι πιθανό να αναπαράγει το ίδιο μοτίβο διακρίσεων³⁰.

Η τεχνητή νοημοσύνη, λοιπόν, θεωρείται προκατειλημμένη όταν παράγει αποτελέσματα που ενισχύουν και διαιωνίζουν στερεότυπα που βλάπτουν συγκεκριμένες ομάδες, ενώ είναι δίκαιη όταν κάνει προβλέψεις ή καταλήγει σε αποτελέσματα που δεν κάνουν (αδικαιολόγητες) διακρίσεις ή ευνοούν κάποια συγκεκριμένη ομάδα³¹.

Χαρακτηριστικά περιστατικά αλγοριθμικής μεροληψίας έλαβαν ευρύτατης δημοσιότητας στον τύπο. Το 2015 η Amazon διαπίστωσε ότι ο αλγόριθμός της που χρησιμοποιήθηκε για την πρόσληψη υπαλλήλων βρέθηκε να είναι προκατειλημμένος σε βάρος των γυναικών παρόλο που αυτές ήταν ισάξιες με τους άντρες, βαθμολογώντας τα βιογραφικά με μη ουδέτερο ως προς το φύλο τρόπο. Το 2019 η πιστωτική κάρτα της Apple, στην οποία το πιστωτικό όριο προσδιορίζονταν βάσει

³⁰ICO, What about fairness, bias and discrimination?, <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-about-fairness-bias-and-discrimination/>, τ.π. 19.11.2024.

³¹ Needle F. (06.06.2023), What is AI bias? [+ Data], <https://blog.hubspot.com/marketing/ai-bias>, τ.π. 28.03.2024.

αλγορίθμου, κατηγορήθηκε ως «σεξιστική» διότι παρείχε σημαντικά χαμηλότερο πιστωτικό όριο στις γυναίκες παρόλο που είχαν παρόμοια οικονομική κατάσταση με τους άντρες³². Η χορήγηση χαμηλότερου πιστωτικού ορίου ή η απόρριψη μιας αίτησης εργασίας συνεπάγονται σοβαρές οικονομικές επιπτώσεις για τους αιτούντες ενώ η χρήση τέτοιου είδους μοντέλων μπορεί να αυξήσει την οικονομική ανισότητα μεταξύ ανδρών και γυναικών. Το 2015 το σύστημα αναγνώρισης προσώπου της Google βρέθηκε να επισημαίνει (image – labeling technology) τις φωτογραφίες έγχρωμων ανθρώπων ως «γορίλλες»³³. Το 2016 η εφημερίδα Propublica έφερε στο φως την υπόθεση αλγοριθμικής μεροληψίας σε βάρος έγχρωμων παραβατών από τον αλγόριθμο COMPAS, ο οποίος χρησιμοποιείται στα δικαστικά συστήματα των ΗΠΑ για την πρόβλεψη της πιθανότητας υποτροπής των κατηγορουμένων³⁴. Το 2020 δημοσιεύθηκε μελέτη η οποία διαπίστωσε ότι τα συστήματα ιατρικών διαγνώσεων με τη βοήθεια υπολογιστή (CAD) τείνουν να έχουν χαμηλότερη ακρίβεια κατά τη διάγνωση υποεκπροσωπούμενων ομάδων, εάν τα σύνολα δεδομένων εκπαίδευσης δεν είναι ισορροπημένα ως προς το φύλο³⁵. Το 2019 η Ολλανδική Αρχή Προστασίας Δεδομένων έκρινε ότι η φορολογική αρχή της Ολλανδίας χρησιμοποίησε ένα σύστημα με την ονομασία SyRI για την ανίχνευση απάτης σχετικά με επιδόματα το οποίο μεροληπτούσε σε βάρος ατόμων διπλής εθνικότητας ή συγκεκριμένης εθνικής καταγωγής βάζοντας στο στόχαστρο γειτονιές όπου ζούσαν άνθρωποι με χαμηλό εισόδημα ή μειονότητες. Το σύστημα αυτό έκρινε ως «ύποπτους για απάτη» 26.000 γονείς κατηγορώντας τους άδικα και λανθασμένα ότι εξαπάτησαν το σύστημα απόδοσης επιδομάτων επιμέλειας ανηλίκων, αναγκάζοντας 10.000 από αυτούς να

³² Apple's 'sexist' credit card investigated by US regulator, (11.11.2019), BBC.com, <https://www.bbc.com/news/business-50365609>, τ.π. 08.08.2024

³³ James V., (12/01/2018), Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech, <https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai>, τ.π. 07.04.2024.

³⁴ Real-life Examples of Discriminating Artificial Intelligence (χ.η.), <https://datatron.com/real-life-examples-of-discriminating-artificial-intelligence/>, τ.π. 27.03.2024.

³⁵ Nidumolu, K.(01.05.2023), Exploring AI Bias in Healthcare: From Gathering Data to Implementation, <https://www.linkedin.com/pulse/exploring-ai-bias-healthcare-from-gathering-data-krishan-dev>, τ.π. 27.03.2024.

επιστρέφουν τα χρήματα των επιδομάτων προκαλώντας οικονομικά και προσωπικά προβλήματα στις οικογένειες αυτές, την αυτοκτονία ενός γονέα³⁶ αλλά και την παραίτηση του πρωθυπουργού της χώρας τον Ιανουάριο 2021 λόγω του σκανδάλου που προκλήθηκε.

Η μαθηματικός και συγγραφέας Cathy O'Neil μάλιστα στο βιβλίο της «Weapons of Math Destruction³⁷» χαρακτηρίζει τα μεροληπτικά μοντέλα ως όπλα μαθηματικής καταστροφής παραφράζοντας τη φράση όπλα μαζικής καταστροφής διότι χρησιμοποιούνται σε διάφορους τομείς (ασφάλειες, προσλήψεις, εκπαίδευση, αστυνόμευση) κατά τρόπο που ενισχύει τις ανισότητες, τιμωρώντας τους φτωχούς και τους καταπιεσμένους της κοινωνίας, κάνοντας τους πλούσιους πλουσιότερους. Το γεγονός ότι οι αλγόριθμοι μπορούν να λάβουν αποφάσεις σε μεγάλη κλίμακα τους καθιστά πιο επικίνδυνους από τους ανθρώπους που μεροληπτούν. Ακολουθώς θα εξηγήσουμε πώς επιφλοχωρεί η μεροληψία στους αλγορίθμους.

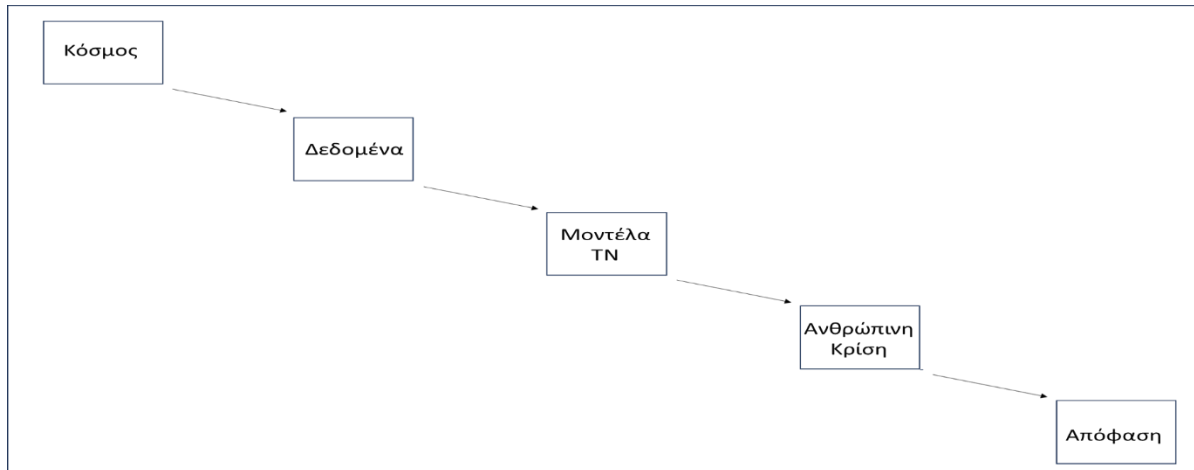
2.2. Οι πηγές του προβλήματος της αλγοριθμικής μεροληψίας

Τα βασικά σημεία όπου μπορεί να εντοπιστεί η μεροληψία στη διαδικασία λήψης αποφάσεων με τη χρήση της τεχνητής νοημοσύνης είναι: α) τα δεδομένα, β) η διαδικασία της μοντελοποίησης και γ) η τελική ανθρώπινη κρίση³⁸. Τα τρία στάδια παρουσιάζονται στο παρακάτω σχήμα:

³⁶ Σ. Τάσσης, Μπορεί ο αλγόριθμος να είναι ηθικός; Τεχνητή νοημοσύνη και ηθικοί κανόνες μια παλιά ιστορία από το μέλλον, σε συλλογικό έργο Μήτρου Λ. Μπορεί ο αλγόριθμος... να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, Πανεπιστημιακές εκδόσεις Κρήτης, Ηράκλειο Κρήτης, 2023, σελ. 53.

³⁷ O'Neil C., Weapons of math destruction, How big data increases inequality and threatens democracy, Penguin Books Ltd, 2017.

³⁸ Reagan M. (25.03.2021), Understanding Bias and Fairness in AI Systems, An illustrated introduction to some of the basic concepts of a crucial problem, published in Towards Data Science, <https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3>, τ.π. 28.03.2024.



Στάδια TN στα οποία εισχωρεί η μεροληψία³⁹

Ακολούθως θα αναλύσουμε πώς εισχωρεί η μεροληψία σε αυτά τα στάδια.

1^ο Στάδιο: Δεδομένα

Η μεροληψία στα δεδομένα μπορεί να λαμβάνει διάφορες μορφές οι σημαντικότερες από τις οποίες είναι η ιστορική μεροληψία, η μεροληψία εκπροσώπησης και η μεροληψία μέτρησης.

A) Ιστορική μεροληψία

Τα ιστορικά δεδομένα ή αλλιώς δεδομένα εκπαίδευσης των αλγορίθμων προέρχονται από τον πραγματικό κόσμο, στον οποίο δυστυχώς υπάρχουν προκαταλήψεις, στερεότυπα και ανισότητες. Οι μέθοδοι τεχνητής νοημοσύνης βασίζονται στα δεδομένα αυτά και ως εκ τούτου είναι φυσικό τα αποτελέσματά τους σε ότι αφορά τη λήψη αποφάσεων να είναι και αυτά ως ένα βαθμό μεροληπτικά, αφού αντανακλούν την ισχύουσα πραγματικότητα στην κοινωνία. Ειδικότερα, μέχρι τα μέσα του 20ού αιώνα, οι γυναίκες συχνά περιοριζόνταν σε ρόλους όπως γραμματείς, δασκάλες, ή νοσοκόμες, ενώ είχαν περιορισμένες ευκαιρίες εκπαίδευσης και επαγγελματικής εξέλιξης σε σχέση με τους άνδρες. Οι γυναίκες μέχρι και σήμερα δύσκολα

³⁹ Reagan M., ο.π.

καταλαμβάνουν θέσεις ευθύνης και θέσεις υψηλών αποδοχών. Στις μέρες μας, παρόλο που έχει σημειωθεί αξιοσημείωτη πρόοδος, μέσω εκπαιδευτικών μεταρρυθμίσεων, κοινωνικών προγραμμάτων και πολιτικών ισότητας, οι ανισότητες και κοινωνικές αδικίες, δυστυχώς, συνεχίζουν να υπάρχουν.

Οι ανισότητες αυτές, λοιπόν, καθρεπτίζονται και στις αλγοριθμικές αποφάσεις. Έτσι, χαρακτηριστικά ένας αλγόριθμος διαλογής υποψηφίων για την εισαγωγή σε ιατρική σχολή της Αγγλίας, βρέθηκε ότι κατέταξε τους λευκούς άντρες υποψήφιους υψηλότερα όταν εξέταζε νέες αιτήσεις υποψηφίων, εξαιτίας του ότι είχε εκπαιδευτεί σε παρελθοντικά ιστορικά δεδομένα λευκών ανδρών που κατά κύριο λόγο γίνονταν δεκτοί στο παρελθόν στη συγκεκριμένη ιατρική σχολή⁴⁰. Όμοια και στην περίπτωση του αλγόριθμου προσλήψεων της Amazon⁴¹ ο αλγόριθμος είχε εκπαιδευτεί σε βιογραφικά που υποβλήθηκαν στην Amazon για μια περίοδο δέκα ετών, τα οποία προέρχονταν κυρίως από άνδρες, αντανakλώντας την ανδροκρατούμενη βιομηχανία της τεχνολογίας και έτσι έμαθε να προτιμάει υποψήφιους που έχουν παρόμοια χαρακτηριστικά με αυτούς που έχουν ήδη προσληφθεί, άρα να προτιμάει τους άντρες.

B) Μεροληψία εκπροσώπησης

Η μεροληψία αυτή σχετίζεται με τον τρόπο που γίνεται η δειγματοληψία από έναν πληθυσμό για να δημιουργηθεί ένα σύνολο δεδομένων εκπαίδευσης, τα οποία θα χρησιμοποιήσει ο αλγόριθμος σαν πρώτη ύλη για να μάθει ώστε να λάβει εν συνεχεία τις αποφάσεις. Η μεροληψία επέρχεται όταν στα δεδομένα εκπαίδευσης υποεκπροσωπούνται κάποιες ομάδες του πληθυσμού. Χαρακτηριστική περίπτωση είναι η ανακάλυψη που έκαναν επιστήμονες δεδομένων στο MIT ότι οι αλγόριθμοι αναγνώρισης προσώπου είχαν υψηλότερα ποσοστά λάθους όταν ανέλυαν πρόσωπα

⁴⁰ Hacker, P. ο.π.

⁴⁰ Information Commissioner's Office (ICO), What about fairness, bias and discrimination?, <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-about-fairness-bias-and-discrimination/>, τ.π. 28.03.2024,

⁴¹ Albaroudi E, Mansouri T, Alameer A. A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. AI. 2024; 5(1):383-404. <https://doi.org/10.3390/ai5010019>

μειονοτήτων, πιθανώς λόγω του ότι τροφοδοτούνταν κυρίως με λευκά ανδρικά πρόσωπα κατά τη διάρκεια της εκπαίδευσης του αλγόριθμου⁴².

Σύμφωνα με σχετική έκθεση του Ερευνητικού Ινστιτούτου AI Now του Πανεπιστημίου της Νέας Υόρκης το οποίο επικεντρώνεται στις κοινωνικές επιπτώσεις της τεχνητής νοημοσύνης, η ανισορροπία μεταξύ των φύλων του εργατικού δυναμικού στις μεγάλες εταιρείες τεχνολογίας όπως η Google, το Facebook και η Microsoft βοηθά στη διαίωση της μεροληψίας στην τεχνητή νοημοσύνη και στην ενίσχυση μιας στενής αντίληψης της έννοιας «κανονικός άνθρωπος», ώστε δημιουργείται μια «κρίση διαφορετικότητας» και η προκατάληψη διεισδύει σε προϊόντα όπως τα προγράμματα αναγνώρισης προσώπου. Οι επιστήμονες που ασχολούνται με την τεχνητή νοημοσύνη και ειδικότερα αυτοί που ασχολούνται με την επιλογή των ιστορικών δεδομένων προς εκπαίδευση μπορούν να προσθέσουν προκατάληψη στο σύστημα βασιζόμενοι σε σύνολα δεδομένων που δεν αντικατοπτρίζουν με ακρίβεια τον κόσμο, όπως βασιζόμενοι σε εικόνες προσώπων που περιλαμβάνουν πολύ λίγους έγχρωμους ανθρώπους⁴³.

Γ) Μεροληψία μέτρησης

Η μεροληψία μέτρησης μπορεί να επέλθει κατά τον ορισμό των χαρακτηριστικών – μεταβλητών του πληθυσμού που θα συλλεχθούν (πχ. επίπεδο εκπαίδευσης) ή κατά τη δημιουργία της μεταβλητής στόχου – label (πχ. καλή/κακή απόδοση εργαζομένου). Τα ιστορικά δεδομένα που είναι εύκολα προσβάσιμα είναι συνήθως παρεμφερή με τις μεταβλητές που μας ενδιαφέρουν, ωστόσο χρησιμοποιώντας τέτοια παρεμφερή ιστορικά δεδομένα μεροληπτούμε στη μέτρηση⁴⁴.

⁴² Yurushkin M. (χ.η.), How Can Synthetic Data Solve the AI Bias Problem?, <https://broutonlab.com/blog/ai-bias-solved-with-synthetic-data-generation>, τ.π. 28.03.2024.

⁴³ Picchi A (17.04.2019), How tech's white male workforce feeds bias into AI, <https://www.cbsnews.com/news/ai-bias-problem-techs-white-male-workforce/>, τ.π. 28.03.2024.

⁴⁴ Reagan M., ο.π.

2^ο Στάδιο: Μοντελοποίηση

Ακόμα κι αν έχουμε τα ιδανικά δεδομένα χωρίς μεροληψία οι διαδικασίες μοντελοποίησης μπορούν να εισάγουν μεροληψία. Οι πιο σημαντικές περιπτώσεις εισαγωγής μεροληψίας κατά τη μοντελοποίηση είναι η μεροληψία που σχετίζεται με την αξιολόγηση του μοντέλου και η μεροληψία κατά τον συνδιασμό και ομαδοποίηση των δεδομένων.

A) Μεροληψία αξιολόγησης

Η μεροληψία αξιολόγησης εμφανίζεται όταν η διαδικασία αξιολόγησης του μοντέλου αποτυγχάνει να εντοπίσει τη μεροληψία στα δεδομένα εκπαίδευσης⁴⁵. Η μεροληψία αξιολόγησης συμβαίνει όταν ένας αλγόριθμος αξιολογείται με δεδομένα που δεν αντιπροσωπεύουν επαρκώς όλες τις ομάδες ή περιπτώσεις στις οποίες θα κληθεί να εφαρμοστεί το μοντέλο στην πράξη. Έτσι, η αξιολόγηση δείχνει καλές επιδόσεις και υψηλή ακρίβεια του μοντέλου, αλλά ο αλγόριθμος στην πραγματικότητα δεν θα αποδίδει εξίσου καλά για κάποιες ομάδες.

B) Μεροληψία ομαδοποίησης – συγκέντωσης και συνδυασμού

Η μεροληψία συγκέντωσης αναφέρεται στη μεροληψία που προκύπτει όταν δεδομένα από διαφορετικές ομάδες συγκεντρώνονται και αναλύονται ως ενιαία ομάδα, χωρίς να λαμβάνονται υπόψη οι διαφορές τους⁴⁶. Ειδικότερα, υπάρχουν πολλές εφαρμογές τεχνητής νοημοσύνης στις οποίες ο πληθυσμός για τον οποίο θέλουμε να βγάλουμε μια απόφαση είναι ανομοιογενής και επομένως ένας μόνο αλγόριθμος κατά πάσα πιθανότητα δεν θα οδηγήσει σε ασφαλές συμπέρασμα για όλες τις υποομάδες που έχουν διαφορετικά χαρακτηριστικά. Για παράδειγμα οι εργαζόμενοι αποτελούνται από άντρες και γυναίκες, με τις τελευταίες να έχουν μεγαλύτερα κενά στο βιογραφικό τους λόγω εγκυμοσύνης και ανατροφής παιδιών. Αν ένας αλγόριθμος χρησιμοποιήσει το χρονικό διάστημα κενού στο βιογραφικό σαν χαρακτηριστικό

⁴⁵ Reagan M., ο.π.

⁴⁶ Reagan M., ο.π.

κρίσιμο για τη μη πρόσληψης εργαζομένου, τότε ο αλγόριθμος μεροληπτεί σε βάρος των γυναικών, οι οποίες αναλαμβάνουν κατά κύριο λόγο την ανατροφή των τέκνων λαμβάνοντας άδειες προστασίας της μητρότητας.

3^ο Στάδιο: Μεροληψία στην τελική ανθρώπινη κρίση

Ακόμα κι αν ο αλγόριθμος κάνει σωστές προβλέψεις, η πιστή εφαρμογή των οποίων θα οδηγούσε σε σωστές αποφάσεις ο λήπτης της απόφασης θα μπορούσε να αγνοήσει τη σύσταση του αλγορίθμου⁴⁷. Έτσι ένα στέλεχος στο τμήμα ανθρώπινου δυναμικού θα μπορούσε να παραβλέψει τη σωστή πρόβλεψη του αλγορίθμου για πρόσληψη μιας γυναίκας υποψήφιας, η οποία έχει όλα τα απαραίτητα προσόντα για να αποδώσει καλά στη συγκεκριμένη θέση εργασίας όπως εκπαίδευση, εμπειρία, προσωπικότητα κ.λ.π. και παρόλαυτα να την αποκλείσει διότι κατά την προσωπική του μεροληπτική κρίση η γυναίκα υποψήφια θα απουσίαζε για μεγάλο χρονικό διάστημα λόγω πιθανής τεκνοποίησης γεγονός το οποίο θα την καθιστούσε κατά την άποψή του λιγότερο αποδοτική στη συγκεκριμένη θέση. Η μεροληψία αυτή δυστυχώς δεν μπορεί να προβλεφθεί ή αποκλειστεί ανεξάρτητα από τα προηγούμενα στάδια που αναφέραμε.

2.3. Το πρόβλημα της έμμεσης μεροληψίας (Proxy Discrimination)

Η μη χρήση ευαίσθητων χαρακτηριστικών (παραμέτρων) στα μοντέλα μηχανικής μάθησης δεν εγγυάται την εξάλειψη της μεροληψίας στις αλγοριθμικές συστάσεις καθώς υπάρχει ο κίνδυνος της έμμεσης μεροληψίας (Proxy discrimination). Έμμεση μεροληψία στην τεχνητή νοημοσύνη είναι το φαινόμενο όπου οι αλγόριθμοι λαμβάνουν αποφάσεις με βάση χαρακτηριστικά ή δεδομένα που αν και δεν είναι προστατευόμενα, λειτουργούν ως έμμεσοι δείκτες ευαίσθητων δεδομένων, όπως το φύλο, η φυλή, η ηλικία κ.λ.π. (proxies - χαρακτηριστικά υποκατάστασης). Οι διακρίσεις που προκύπτουν από τη χρήση υποκατάστατων χαρακτηριστικών είναι πιο δύσκολο να αποτραπούν, καθώς υπάρχει ένας δυνητικά απεριόριστος αριθμός υποκατάστατων μεταβλητών.

⁴⁷ Reagan M., ο.π.

Όπως εξηγήσαμε στη μηχανική μάθηση ο αλγόριθμος αυτόνομα εντοπίζει από τα περισσότερα χαρακτηριστικά με τα οποία έχει τροφοδοτηθεί εκείνα που θεωρεί ότι σχετίζονται περισσότερο με τον επιθυμητό στόχο (πχ. καλή απόδοση στην εργασία, πιθανότητα αποπληρωμής δανείου). Τα συστήματα τεχνητής νοημοσύνης δεν είναι σε θέση να διακρίνουν μεταξύ αιτιοτήτων και συσχετισμών, με αποτέλεσμα να προβαίνουν σε κατηγοριοποιήσεις επί τη βάσει συσχετισμών χωρίς να συνδέονται τα χαρακτηριστικά που χρησιμοποιούνται για την κατηγοριοποίηση απαραίτητα αιτιωδώς με μια κατηγορία ή μια συμπεριφορά⁴⁸. Αυτό σημαίνει ότι τα συστήματα τεχνητής νοημοσύνης μπορεί να βρουν μια σύνδεση μεταξύ δύο χαρακτηριστικών στα δεδομένα τους, χωρίς να κατανοούν αν το ένα προκαλεί το άλλο ή αν πρόκειται απλώς για τυχαία συνύπαρξη.

Έτσι, στην περίπτωση που ένας αλγόριθμος για την αυτοματοποιημένη πρόσληψη εργαζομένων καταλήγει στο συμπέρασμα ότι οι υποψήφιοι οι οποίοι έχουν μεγάλο διάστημα απουσίας από την εργασία τους (φαινομενικά ουδέτερη μεταβλητή: «συνολικός χρόνος απουσίας από εργασία») θα έχουν χαμηλότερη απόδοση και επομένως δεν θα πρέπει προσληφθούν (μεταβλητή στόχος: «πρόσληψη ή μη πρόσληψη»), καταλήγει εν τέλει σε έμμεση διάκριση σε βάρος των γυναικών, οι οποίες κατά το σύνηθες έχουν κενά εργασίας λόγω άδειας ανατροφής τέκνων. Φαινομενικά ουδέτερα χαρακτηριστικά είναι και η απόσταση της κατοικίας του υποψήφιου εργαζόμενου από την εργασία ή οι σπουδές σε φημισμένα πανεπιστήμια, κριτήρια τα οποία, όμως, μπορεί τελικά να σχετίζονται με την κοινωνικοοικονομική κατάσταση των υποψηφίων (πχ. περιοχές με υψηλή συγκέντρωση συγκεκριμένων εθνοτικών ομάδων ή ομάδων χαμηλών εισοδημάτων, πανεπιστήμια με πολύ ακριβά δίδακτρα). Στον μη αλγοριθμικό κόσμο ως τέτοια ύποπτα κριτήρια για την επιλογή προς πρόσληψη έχουν χαρακτηριστεί από τη νομολογία η πρόβλεψη ελάχιστου ύψους, η απαίτηση για προηγούμενη πλήρη απασχόληση, η απαίτηση για

⁴⁸ Μήτρου Λ., Μπορεί ο αλγόριθμος να διοικεί; σε συλλογικό έργο Μήτρου Λ. Μπορεί ο αλγόριθμος... να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, Πανεπιστημιακές εκδόσεις Κρήτης, Ηράκλειο Κρήτης, 2023, σελ. 278.

προσαρμοστικότητα σε μεταβαλλόμενα ωράρια και τρόπους εργασίας και η αρχαιότητα⁴⁹. Η έμμεση μεροληψία είναι ένα σημαντικό ζήτημα στην ανάπτυξη μοντέλων μηχανικής μάθησης και είναι απαραίτητο να εντοπίζεται και να διορθώνεται ώστε να διασφαλίζεται ότι τα μοντέλα είναι δίκαια και αντικειμενικά.

ΚΕΦΑΛΑΙΟ 3 – Η ΔΙΚΑΙΟΣΥΝΗ ΚΑΙ ΙΣΟΤΗΤΑ – ΜΗ ΔΙΑΚΡΙΣΗ ΣΤΗΝ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

3.1. Η ηθική αρχή της δικαιοσύνης (fairness) στην τεχνητή νοημοσύνη

Το πρόβλημα της αλγοριθμικής μεροληψίας ανήκει στους ηθικούς προβληματισμούς που εγείρει η χρήση της τεχνητής νοημοσύνης. Η ηθική της τεχνητής νοημοσύνης (Artificial Intelligence Ethics) είναι ο τομέας που μελετά τις ηθικές αρχές, τους κανονισμούς και κατευθυντήριες γραμμές που διέπουν την ανάπτυξη και τη χρήση τεχνολογιών τεχνητής νοημοσύνης. Οργανισμοί αλλά και εταιρείες κολοσσοί της τεχνολογίας ασχολούνται με το ζήτημα της ηθικής στην τεχνητή νοημοσύνη και την κατάρτιση πολιτικών για τη ρύθμιση της, όπως η Ευρωπαϊκή Επιτροπή⁵⁰, ο ΟΟΣΑ⁵¹, η UNESCO⁵², οι Amazon, Google, Facebook, IBM, Microsoft, Apple⁵³. Στα περισσότερα κείμενα υπάρχει σύμπνοια για το ποιες είναι οι ηθικές αρχές που θα πρέπει να τηρούνται προκειμένου να αντιμετωπίζονται οι ηθικές προκλήσεις που εγείρει η τεχνητή νοημοσύνη ώστε η τελευταία να λειτουργεί ηθικά. Έτσι, με δεδομένο ότι ο σχεδιασμός και η ανάπτυξη συστημάτων τεχνητής νοημοσύνης μπορούν να οδηγήσουν σε προσβολή δικαιωμάτων και αξιών γίνεται αναφορά στην ανάγκη

⁴⁹ Παπαγεωργίου Ι., Συστήματα τεχνητής νοημοσύνης και διακριτική μεταχείριση κατά το προσυμβατικό στάδιο σύναψης της σύμβασης εργασίας, ΕφΑΔΠολΔ, 10/2022, σελ. 1125 – 1140.

⁵⁰ Κατευθυντήριες γραμμές δεοντολογίας για αξιόπιστη τεχνητή νοημοσύνη της Ανεξάρτητης Ομάδας Εμπειρογνομώνων Υψηλού Επιπέδου για την τεχνητή νοημοσύνη η οποία συστήθηκε από την Ευρωπαϊκή Επιτροπή (2019), <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>, τ.π. 27.03.2024.

⁵¹ Ιστοσελίδα OECD AI Policy Observatory, <https://oecd.ai/en/>, τ.π. 27.03.2024

⁵² Recommendation on the Ethics of Artificial Intelligence, UNESCO, 2021 <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>, τ.π. 27.03.2024

⁵³ Fiegeman S. (28/09/2016), «Facebook, Google, Amazon create group to ease AI concerns», money.cnn.com, <https://money.cnn.com/2016/09/28/technology/partnership-on-ai/>, τ.π. 27.03.2024.

τήρησης των αρχών της ελευθερίας, ιδιωτικότητας, δικαιοσύνης, διαφάνειας, λογοδοσίας και ευημερίας των ατόμων, της κοινωνίας και του περιβάλλοντος. Πολλές από αυτές τις αξίες αντιστοιχούν στα ανθρώπινα δικαιώματα όπως κατοχυρώνονται στο διεθνές και εθνικό δίκαιο και στις γενικές αξίες που ενστερνίζονται οι πολιτισμένες χώρες⁵⁴.

Με δεδομένο ότι έχει αναπτυχθεί σε ευρωπαϊκό επίπεδο θα αναφερθούμε στην έννοια της αρχής της δικαιοσύνης όπως αποτυπώνεται στις κατευθυντήριες γραμμές της συσταθείσας από την Ευρωπαϊκή Επιτροπή Ανεξάρτητης Ομάδας εμπειρογνομόνων υψηλού επιπέδου για την τεχνητή νοημοσύνη⁵⁵.

Σύμφωνα, λοιπόν, με την αρχή της δικαιοσύνης τα συστήματα τεχνητής νοημοσύνης θα πρέπει να αναπτύσσονται και να χρησιμοποιούνται με δίκαιο τρόπο. Η δικαιοσύνη έχει ουσιαστική και διαδικαστική διάσταση. Η ουσιαστική διάσταση συνεπάγεται ότι τα βάρη και οι ωφέλειες θα πρέπει να κατανέμονται ισότιμα και δίκαια και τα άτομα και οι ομάδες δεν θα πρέπει να υφίστανται αθέμιτη μεροληψία, διακρίσεις και στιγματισμό. Η χρήση συστημάτων τεχνητής νοημοσύνης δεν θα πρέπει ποτέ να οδηγεί στην εξαπάτηση των τελικών χρηστών ή να περιορίζει την ελευθερία επιλογής τους. Επιπλέον, οι επαγγελματίες του τομέα της τεχνητής νοημοσύνης θα πρέπει να σέβονται την αρχή της αναλογικότητας μεταξύ μέσων και σκοπών. Η διαδικαστική διάσταση της δικαιοσύνης συνεπάγεται τη δυνατότητα αμφισβήτησης και αποτελεσματικής έννομης προστασίας έναντι αποφάσεων που λαμβάνονται από συστήματα τεχνητής νοημοσύνης και από τους ανθρώπους που τα χειρίζονται. Για το λόγο αυτό η οντότητα που είναι υπεύθυνη για την απόφαση θα πρέπει να είναι προσδιορισμένη και θα πρέπει να επεξηγούνται οι διαδικασίες λήψης αποφάσεων. Οι Κατευθυντήριες Γραμμές Δεοντολογίας για Αξιόπιστη Τεχνητή Νοημοσύνη αν και

⁵⁴Brey, P., Dainow, B. (2023) Ethics by design for artificial intelligence. AI Ethics, <https://link.springer.com/article/10.1007/s43681-023-00330-4>, τ.π. 28.03.2024.

⁵⁵Κατευθυντήριες γραμμές δεοντολογίας για αξιόπιστη τεχνητή νοημοσύνη της Ανεξάρτητης Ομάδας Εμπειρογνομόνων Υψηλού Επιπέδου για την τεχνητή νοημοσύνη συσταθείσας από την Ευρωπαϊκή Επιτροπή (2019), διαθέσιμο στον ακόλουθο σύνδεσμο: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>, τ.π. 27.03.2024

δεν είναι νομικά δεσμευτικές αποτυπώνουν τη λογική των νομοθετικών κειμένων που ακολουθούν, όπως ο Κανονισμός για την τεχνητή νοημοσύνη, ο οποίος αποτελεί τη μετάβαση από την ηθική στο δίκαιο⁵⁶.

3.2. Η ενσωμάτωση της δικαιοσύνης ήδη από τον σχεδιασμό της τεχνητής νοημοσύνης (Ethics by Design)

Η έννοια της προστασίας της ιδιωτικότητας και των προσωπικών δεδομένων ήδη από τον σχεδιασμό των πράξεων επεξεργασίας (privacy by design) μας είναι ήδη γνώριμη από τον Γενικό Κανονισμό Προστασίας Δεδομένων (αρ. 25 και αιτ. σκέψη 78 ΓΚΠΔ). Πλέον, με την «αναγέννηση» της τεχνητής νοημοσύνης έχει προκύψει η έννοια της «Ηθικής τεχνητής νοημοσύνης από τον σχεδιασμό» (Ethics by Design). Στόχος της «Ηθικής τεχνητής νοημοσύνης από τον σχεδιασμό» είναι η ενσωμάτωση των ηθικών αρχών κατά τη διαδικασία ανάπτυξης της εκάστοτε εφαρμογής της τεχνητής νοημοσύνης ώστε τα ηθικά ζητήματα που γεννώνται να αντιμετωπίζονται όσο το δυνατόν νωρίτερα.

Σχετική είναι η από 25.11.2021, κατόπιν ανάθεσης της Ευρωπαϊκής Επιτροπής, μελέτη ειδικών επιστημόνων με τίτλο “Ethics by Design and Ethics of Use Approaches for Artificial Intelligence” («Προσεγγίσεις της Ηθικής από τον σχεδιασμό και της Ηθικής Χρήσης για την τεχνητή νοημοσύνη»), η οποία αποσκοπεί στην ευαισθητοποίηση της επιστημονικής κοινότητας και όσων επωφελούνται από έργα που αξιοποιούν την τεχνητή νοημοσύνη. Η μελέτη αυτή στηρίχθηκε στις ως άνω Κατευθυντήριες για αξιόπιστη τεχνητή νοημοσύνη της ανεξάρτητης ομάδας εμπειρογνομόνων υψηλού επιπέδου για την τεχνητή νοημοσύνη και στα χρηματοδοτούμενα από την ΕΕ έργα SHERPA και SIENNA ⁵⁷.

⁵⁶ Μήτρου Λ., Μερικές εισαγωγικές σκέψεις, σε συλλογικό έργο Μήτρου Λ. Μπορεί ο αλγόριθμος... να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, Πανεπιστημιακές εκδόσεις Κρήτης, Ηράκλειο Κρήτης, 2023, σελ. 31-32.

⁵⁷ European Commission, Ethics By Design and Ethics of Use Approaches for Artificial Intelligence (1.0), European Commission - DG Research and Innovation, 25 November 2021, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf, Guidelines for Trustworthy AI

Αναφορικά με την αρχή της Δικαιοσύνης (Fairness) τονίζεται χαρακτηριστικά στην εν λόγω μελέτη όλοι οι άνθρωποι έχουν δικαίωμα στα ίδια θεμελιώδη δικαιώματα και στις ίδιες ευκαιρίες και δεν πρέπει να υφίστανται διακρίσεις με βάση τις θεμελιώδεις πτυχές της ταυτότητας τους όπως το φύλο, η φυλή, η ηλικία, ο σεξουαλικός προσανατολισμός, η εθνική καταγωγή, η θρησκεία, η υγεία και η αναπηρία. Η διαδικαστική δικαιοσύνη απαιτεί η διαδικασία να μην έχει σχεδιαστεί με τέτοιο τρόπο ώστε να αδικεί μεμονωμένα άτομα ή ομάδες. Η ουσιαστική δικαιοσύνη συνίσταται στην μη υιοθέτηση μοτίβων διάκρισης που επιβαρύνουν αδικαιολόγητα άτομα ή/και ομάδες λόγω των συγκεκριμένων ιδιαιτεροτήτων τους. Η δικαιοσύνη μπορεί να υποστηριχθεί από πολιτικές που προάγουν τη διαφορετικότητα και αξιολογούν θετικά τις διαφορές στους ανθρώπους, συμπεριλαμβανομένων όχι μόνο χαρακτηριστικών όπως το φύλο και τη φυλή, αλλά και τις διαφορετικές προσωπικότητες, τις εμπειρίες, το πολιτιστικό υπόβαθρο, το γνωστικό επίπεδο των ανθρώπων, το στυλ και άλλα χαρακτηριστικά που επηρεάζουν την προσωπική άποψη του καθενός. Η υποστήριξη της διαφορετικότητας σημαίνει υποστήριξη της ποικιλόμορφης σύνθεσης των ομάδων και οργανώσεων⁵⁸. Προκειμένου να τηρηθεί η αρχή της δικαιοσύνης πρέπει να αποφεύγεται η αλγοριθμική μεροληψία. Τα συστήματα τεχνητής νοημοσύνης θα πρέπει να σχεδιάζονται έτσι ώστε να αποφεύγεται η μεροληψία στα δεδομένα εισόδου, κατά την μοντελοποίηση και τον σχεδιασμό των αλγορίθμων. Η αλγοριθμική μεροληψία απαιτεί συγκεκριμένες τεχνικές μετριάσμού. Οι ερευνητικές προτάσεις θα πρέπει να προσδιορίζουν τα βήματα που θα γίνουν ώστε να βεβαιώνεται ότι τα δεδομένα για τα άτομα είναι αντιπροσωπευτικά του πληθυσμού-στόχου και αντικατοπτρίζουν την ποικιλομορφία του ή είναι επαρκώς ουδέτερα. Ακόμη, θα πρέπει να τεκμηριώνουν ρητά τον τρόπο

(<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>), SHERPA (<https://www.project-sherpa.eu>), SIENNA (<https://www.sienna-project.eu>), τ.π. 28.03.2024.

⁵⁸European Commission, Ethics By Design and Ethics of Use Approaches for Artificial Intelligence (1.0), European Commission - DG Research and Innovation, 25 November 2021. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf, Guidelines for Trustworthy AI (<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>), SHERPA (<https://www.project-sherpa.eu>), SIENNA (<https://www.sienna-project.eu>), τ.π. 28.03.2024.

εντοπισμού και αποφυγής της μεροληψίας στα δεδομένα εισόδου και στον σχεδιασμό του αλγορίθμου. Επιπλέον, θα πρέπει να εξασφαλίζεται η καθολική προσβασιμότητα, δηλαδή τα συστήματα τεχνητής νοημοσύνης (όποτε είναι δυνατόν) να σχεδιάζονται έτσι ώστε να μπορούν να χρησιμοποιηθούν από διαφορετικούς χρήστες με διαφορετικές ικανότητες⁵⁹.

Ο Κανονισμός για την τεχνητή νοημοσύνη, στον οποίο θα αναφερθούμε σε ακόλουθο κεφάλαιο, ενστερνίζεται την «Ηθική από τον σχεδιασμό», ιδίως μέσω των άρθρων 9, 10, 13, 14 και 15 όπου θεσπίζονται απαιτήσεις για τη διαχείριση κινδύνου, τη διακυβέρνηση δεδομένων, τη διαφάνεια, την ανθρώπινη εποπτεία και την ακρίβεια των συστημάτων τεχνητής νοημοσύνης, ώστε να διασφαλίζεται ότι τα συστήματα τεχνητής νοημοσύνης σχεδιάζονται και αναπτύσσονται με τρόπο που ευθυγραμμίζεται με τις ηθικές αρχές, προστατεύοντας τα θεμελιώδη δικαιώματα και προάγοντας τη δικαιοσύνη και την λογοδοσία.

Η ηθική από τον σχεδιασμό αποτελεί πλέον χαρακτηριστικό το οποίο τονίζεται κατά τη διαφήμιση προϊόντων τεχνητής νοημοσύνης. Χαρακτηριστικά, όταν κυκλοφόρησε το GPT-4, τον Μάρτιο του 2023, η OpenAI για την προώθηση του προϊόντος στην αγορά υποστήριξε ότι η νέα έκδοση ήταν καλύτερη όσον αφορά την ακρίβεια, την ικανότητα συλλογισμού και τις βαθμολογίες δοκιμών αλλά και «πιο ευθυγραμμισμένη με τις ανθρώπινες αξίες». Καθώς η ευθυγράμμιση των εφαρμογών της τεχνητής νοημοσύνης με τις ηθικές αρχές δεν αποτελεί μόνο ρυθμιστική απαίτηση αλλά και παράγοντα διαφοροποίησης των προϊόντων στην αγορά, οι εταιρείες που θα προσαρμόσουν νωρίς τις διαδικασίες ανάπτυξης των προϊόντων

⁵⁹European Commission, Ethics By Design and Ethics of Use Approaches for Artificial Intelligence (1.0), European Commission - DG Research and Innovation, 25 November 2021. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf, Guidelines for Trustworthy AI (<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>), SHERPA (<https://www.project-sherpa.eu>), SIENNA (<https://www.sienna-project.eu>), τ.π. 28.03.2024.

τους θα αποκτήσουν σημαντικό ανταγωνιστικό πλεονέκτημα κατά την εμπορική προώθηση των προϊόντων τους⁶⁰.

3.3. Η ισότητα - μη διάκριση στο δίκαιο

Η αρχή της ισότητας και της μη διάκρισης αποτελεί θεμελιώδη αρχή που διαπνέει το εθνικό και κοινοτικό δίκαιο. Ειδικότερα, η Ευρωπαϊκή Σύμβαση Δικαιωμάτων του Ανθρώπου (ΕΣΔΑ) προβλέπει στο άρθρο 14 αυτής ότι η χρήση των δικαιωμάτων που κατοχυρώνονται σε αυτήν θα πρέπει εξασφαλίζεται ασχέτως διακρίσεως φύλου, φυλής, χρώματος, γλώσσας, θρησκείας, πολιτικών ή άλλων πεποιθήσεων, εθνικής ή κοινωνικής προέλευσης, συμμετοχής σε εθνική μειονότητα, παρουσίας, γεννήσεως ή άλλης καταστάσεως. Ομοίως, ο Χάρτης των Θεμελιωδών Δικαιωμάτων της ΕΕ τονίζει το απαραβίαστο της ανθρωπίνης αξιοπρέπειας (άρθρο 1), την ισότητα όλων ενώπιον του νόμου (άρθρο 20), την απαγόρευση κάθε μορφής διάκρισης, ιδίως λόγω φύλου, φυλής, χρώματος, εθνοτικής ή κοινωνικής καταγωγής, γενετικών χαρακτηριστικών, γλώσσας, θρησκείας ή πεποιθήσεων, πολιτικών ή άλλων φρονημάτων, ιδιοτητας μέλους εθνικής μειονότητας, παρουσίας, γέννησης, αναπηρίας, ηλικίας ή γενετήσιου προσανατολισμού (άρθρο 21) και την ισότητα των δύο φύλων σε όλους τους τομείς, μεταξύ άλλων στην απασχόληση, την εργασία και τις αποδοχές (άρθρο 23).

Σε εθνικό επίπεδο το Ελληνικό Σύνταγμα προστατεύει την ισότητα και την αξία του ανθρώπου μέσα από διάφορα άρθρα. Έτσι, στο άρθρο 2 παρ. 1 Σ κατοχυρώνεται ο σεβασμός και η προστασία της αξίας του ανθρώπου ως πρωταρχική υποχρέωση της Πολιτείας, στο άρθρο 4 παρ. 1 και 2 Σ η ισότητα όλων των Ελλήνων ενώπιον του νόμου και η ισότητα των Ελλήνων και Ελληνίδων στα δικαιώματα και τις υποχρεώσεις, στο άρθρο 5 παρ. 2 Σ η απόλυτη προστασία της ζωής, της τιμής και της ελευθερίας όλων των ατόμων που βρίσκονται στην ελληνική επικράτεια ασχέτως εθνικότητας, φυλής, γλώσσας και θρησκευτικών ή πολιτικών πεποιθήσεων και στο

⁶⁰ Abernethy J., Candelon F., Evgeniou T., Gupta A., Lostanlen Y., Bring Human Values to AI, Harvard Business Review, March – April 2024 <https://hbr.org/2024/03/bring-human-values-to-ai> , τ.π. 20.02.2024.

άρθρο 22 Σ το δικαίωμα όλων των εργαζομένων σε ίση αμοιβής για παρεχόμενη εργασία ίσης αξίας ανεξάρτητα από το φύλο ή άλλη διάκριση.

Περαιτέρω, σχετικές είναι οι κοινοτικές οδηγίες 2006/54/EK του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 5ης Ιουλίου 2006 για την εφαρμογή της αρχής των ίσων ευκαιριών και της ίσης μεταχείρισης ανδρών και γυναικών σε θέματα εργασίας και απασχόλησης, 2000/43/EK του Συμβουλίου της 29^{ης} Ιουνίου 2000 περί εφαρμογής της αρχής της ίσης μεταχείρισης προσώπων ασχέτως φυλετικής ή εθνικής τους καταγωγής και 2000/78/EK του Συμβουλίου της 27ης Νοεμβρίου 2000 για τη διαμόρφωση γενικού πλαισίου για την ίση μεταχείριση στην απασχόληση και την εργασία.

Οι οδηγίες αυτές ενσωματώθηκαν στην ελληνική έννομη τάξη με τους Ν. 3896/2010 και Ν. 4443/2016. Με βάση τους νόμους αυτούς απαγορεύονται κατά κανόνα οι κάθε μορφής διακρίσεις λόγω φυλής, χρώματος, εθνικής ή εθνικής καταγωγής, γενεαλογικών καταβολών, θρησκευτικών ή άλλων πεποιθήσεων, αναπηρίας ή χρόνιας πάθησης, ηλικίας, οικογενειακής κοινωνικής κατάστασης, σεξουαλικού προσανατολισμού, ταυτότητας ή χαρακτηριστικών ή έκφρασης φύλου στον τομέα της εργασίας και της απασχόλησης. Η απαγόρευση αυτή αφορά: α) τους όρους πρόσβασης στην εργασία και τους όρους υπηρεσιακής και επαγγελματικής εξέλιξης, β) την πρόσβαση στον επαγγελματικό προσανατολισμό και επιμόρφωση, γ) τους όρους και τις συνθήκες εργασίας και απασχόλησης, ιδίως όσον αφορά τις αποδοχές, την απόλυση, την υγεία και την ασφάλεια στην εργασία και σε περίπτωση ανεργίας την επανένταξη και την εκ νέου απασχόληση και δ) τη συμμετοχή σε συνδικαλιστική οργάνωση. Οι διακρίσεις λαμβάνουν τη μορφή της άμεσης ή έμμεσης διάκρισης. Ως «άμεση διάκριση» θεωρείται η εξαιτίας ενός από τα προαναφερθέντα χαρακτηριστικά λιγότερο ευνοϊκή μεταχείριση ενός προσώπου σε σχέση με αυτή που θα τύγχανε ένα άλλο πρόσωπο σε ανάλογη κατάσταση. Έτσι, για παράδειγμα άμεση διάκριση είναι όταν ένας εργοδότης αρνείται να προσλάβει έναν εργαζόμενο μόνο και μόνο επειδή

είναι γυναίκα⁶¹. Από την άλλη, «έμμεση διάκριση» υπάρχει όταν μια εκ πρώτης όψεως ουδέτερη διάταξη, κριτήριο ή πρακτική μπορεί να θέσει πρόσωπα που φέρουν κάποιο από τα προστατευόμενα γνωρίσματα, σε μειονεκτική θέση συγκριτικά με άλλα πρόσωπα. Παράδειγμα, έμμεσης διάκρισης είναι όταν ένας εργοδότης σε αγγελία ζητά να προσλάβει εργαζόμενους που έχουν ύψος πάνω από 1,80 εκ. Το ύψος είναι ένα ουδέτερο μη προστατευόμενο χαρακτηριστικό, όμως, χρησιμοποιώντας αυτό το φαινομενικά ουδέτερο κριτήριο τελικά αποκλείονται έμμεσα οι γυναίκες που από τη φύση τους δεν είναι συνήθως τόσο ψηλές⁶². Επίσης, όταν ένας εργοδότης ζητά να προσλάβει υπαλλήλους με δίπλωμα οδήγησης χωρίς η απαίτηση αυτή να σχετίζεται με τα καθήκοντα της θέσης εμμέσως προβαίνει σε διακριτική μεταχείριση σε βάρος των ατόμων με αναπηρίες που δεν οδηγούν⁶³.

Οι ως άνω διακρίσεις, ωστόσο, σε ορισμένες ρητώς προβλεπόμενες περιπτώσεις κατ' εξαίρεση επιτρέπονται. Έτσι, δεν συνιστά ανεπίτρεπτη διάκριση η διαφορετική μεταχείριση η οποία οφείλεται στα προαναφερθέντα χαρακτηριστικά αν το χαρακτηριστικό λόγω της φύσης συγκεκριμένου επαγγέλματος αποτελεί ουσιαστική επαγγελματική προϋπόθεση και ο σκοπός της διάκρισης είναι θεμιτός και η προϋπόθεση ανάλογη. Ακόμα, δεν συνιστά ανεπίτρεπτη διάκριση η διαφορετική μεταχείριση λόγω ηλικίας εφόσον αυτό προβλέπεται στο νόμο για την εξυπηρέτηση σκοπών της πολιτικής της απασχόλησης, της αγοράς εργασίας και της επαγγελματικής κατάρτισης τα δε μέσα επίτευξης των σκοπών αυτών είναι πρόσφορα και αναγκαία.

Ο Κανονισμός για την Τεχνητή Νοημοσύνη ενώ προβλέπει μέτρα για τον μετριασμό της αλγοριθμικής μεροληψίας δεν καθιστά σαφές σε ποιες περιπτώσεις είναι παράνομη η άνιση μεταχείριση και αφήνει αυτήν την κρίση περί παράνομης

⁶¹Wachter S, Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI, Oxford Internet Institute, University of Oxford online webinar <https://www.oii.ox.ac.uk/news-events/events/why-fairness-cannot-be-automated-bridging-the-gap-between-eu-non-discrimination-law-and-ai/>, τ.π. 16.03.2024.

⁶² Wachter S, ο.π.

⁶³ Οδηγός Ίσης Μεταχείρισης – Εγχειρίδιο για δημοσίους υπαλλήλους από το Συνήγορο του Πολίτη (27.07.2018), <https://old.synigoros.gr/?i=equality.el.ifleaflet>, τ.π. 23.04.2024

διάκρισης στο ευρωπαϊκό δίκαιο που απαγορεύει τις διακρίσεις. Ειδικότερα, σύμφωνα με την αιτιολογική σκέψη 6 του Κανονισμού, είναι ζωτικής σημασίας η τεχνητή νοημοσύνη να αναπτύσσεται με τρόπο σύμφωνο προς τις αξίες της ΕΕ όπως κατοχυρώνονται στο άρθρο 2 της Συνθήκης για την ΕΕ. Στις αξίες αυτές συγκαταλέγεται και το δικαίωμα στην ισότητα και μη διάκριση. Ακόμη, σύμφωνα με το άρθρο 10 παρ. 2 του Κανονισμού τα σύνολα δεδομένων εκπαίδευσης, επικύρωσης και δοκιμής υπόκεινται σε πρακτικές διακυβέρνησης και διαχείρισης δεδομένων, οι οποίες ενδείκνυνται για τον επιδιωκόμενο σκοπό του συστήματος υψηλού κινδύνου και εξετάζονται έναντι πιθανών περιπτώσεων μεροληψίας που μπορεί να οδηγήσουν σε διακρίσεις που απαγορεύονται βάσει του ενωσιακού δικαίου⁶⁴.

3.4. Η Αλγοριθμική Δικαιοσύνη (Algorithmic Fairness)

Η αλγοριθμική δικαιοσύνη (algorithmic fairness) είναι το πεδίο έρευνας που στοχεύει στην κατανόηση και τη διόρθωση των προκαταλήψεων και περιλαμβάνει την έρευνα των αιτιών της μεροληψίας σε δεδομένα και αλγόριθμους, τον καθορισμό και εφαρμογή μετρικών δικαιοσύνης (fairness metrics), την ανάπτυξη μεθοδολογιών συλλογής δεδομένων και μοντελοποίησης με στόχο τη δημιουργία δίκαιων αλγορίθμων και την παροχή συμβουλών σε κυβερνήσεις και εταιρίες σχετικά με τον τρόπο ρύθμισης της μηχανικής μάθησης⁶⁵. Η μέτρηση της δικαιοσύνης στις προβλέψεις, δηλαδή η αξιολόγηση κατά πόσο ένα μοντέλο είναι δίκαιο και όχι άδικο και μεροληπτικό, γίνεται από τους επιστήμονες των δεδομένων οι οποίοι για το σκοπό αυτό χρησιμοποιούν διάφορους ορισμούς της δικαιοσύνης (fairness metrics). Ειδικότερα, αφού χωρίσουν τις ομάδες σε προνομιούχες (π.χ. άνδρες) και μη προνομιούχες (π.χ. γυναίκες) τις συγκρίνουν χρησιμοποιώντας διάφορες μετρήσεις - μετρικές. Για παράδειγμα, σύμφωνα με τον ορισμό των εξισωμένων πιθανοτήτων (equalized odds) απαιτείται τα αληθινά θετικά ποσοστά (π.χ. σωστή χορήγηση

⁶⁴ L. Deck, J. Müller, C. Braun, D. Zipperling, N. Kühn (04.01.2024), Implications of the AI Act for non-discrimination law and algorithmic Fairness, <https://arxiv.org/abs/2403.20089>, τ.π. 14.08.2024

⁶⁵ Conor O'Sullivan (05.03.2021), What is Algorithm Fairness? An introduction to the field that aims at understanding and preventing unfairness in machine learning, Towards Data Science, <https://towardsdatascience.com/what-is-algorithm-fairness-3182e161cf9f>, τ.π. 08.08.2024

δανείου σε όσους μπορούν να αποπληρώσουν, σωστή σύσταση για προαγωγή σε όσους την αξίζουν) και τα ψευδώς θετικά ποσοστά (π.χ. εσφαλμένη χορήγηση δανείου σε κάποιον που δεν μπορεί να αποπληρώσει, σύσταση για προαγωγή σε κάποιον που δεν την αξίζει) των δύο ομάδων να είναι ίσα. Ένα μοντέλο με σημαντικά διαφορετικά ποσοστά θεωρείται άδικο. Άλλοι ορισμοί περιλαμβάνουν τις ίσες ευκαιρίες (equal opportunity), όπου απαιτείται τα αληθινά θετικά ποσοστά να είναι ίδια για όλες τις ομάδες ώστε όλες οι ομάδες να έχουν ίσες ευκαιρίες να λάβουν ένα θετικό αποτέλεσμα εφόσον το αξίζουν⁶⁶ και τη δημογραφική ισότητα (Demographic parity) όπου το ποσοστό των θετικών αποφάσεων πρέπει να είναι το ίδιο για όλες τις ομάδες, ανεξάρτητα από το αν τα άτομα αξίζουν ή όχι την απόφαση (πχ. το ποσοστό των προαγωγών πρέπει να είναι το ίδιο, ανεξάρτητα από το φύλο ακόμη και αν υπάρχουν διαφοροποιήσεις στις επιδόσεις).

Ωστόσο, αν και οι μετρικές δικαιοσύνης είναι πολύ χρήσιμα εργαλεία σήμερα δεν υπάρχει ένα καθολικά αποδεκτό πρότυπο για μετρικές δικαιοσύνης, γεγονός το οποίο δημιουργεί προκλήσεις στη διασφάλιση συνεπούς νομικής συμμόρφωσης σε διαφορετικά συστήματα και δικαιοδοσίες τεχνητής νοημοσύνης. Το πρόβλημα με τις μετρικές δικαιοσύνης στους αλγορίθμους είναι ότι η έννοια της δικαιοσύνης είναι πολύπλευρη και μπορεί να ερμηνευτεί με διάφορους τρόπους, κάτι που καθιστά δύσκολη την εφαρμογή της σε πρακτικό επίπεδο⁶⁷. Η δικαιοσύνη, δεν είναι μια εύκολα μετρήσιμη αξία που μπορεί να εκφραστεί μέσω των μαθηματικών και της στατιστικής αλλά μια πανανθρώπινη αξία που κρίνεται σύμφωνα με τα συγκεκριμένα κάθε φορά υπό κρίση πραγματικά περιστατικά⁶⁸. Οι μετρικές δικαιοσύνης, συνεπώς, πρέπει να χρησιμοποιούνται με προσοχή και σε συνδυασμό με

⁶⁶Conor O'Sullivan (05.03.2021), What is Algorithm Fairness?An introduction to the field that aims at understanding and preventing unfairness in machine learning, Towards Data Science, <https://towardsdatascience.com/what-is-algorithm-fairness-3182e161cf9f>, τ.π. 08.08.2024

⁶⁷ Alvarez, J.M., Colmenarejo, A.B., Elobaid, A. et al. Policy advice and best practices on bias and fairness in AI. Ethics Inf Technol 26, 31 (2024). <https://doi.org/10.1007/s10676-024-09746-w>, τ.π. 28.08.2024

⁶⁸ Koutsoviti Koumeri L., Legast M., Yousefi Y., Vanhoof K, Legay A, and Schommer C. 2023. Compatibility of FairnessMetrics with EU Non-Discrimination Laws: Demographic Parity & Conditional Demographic Disparity. In Proceedings of ACM Conference (Conference'17). ACM, New York, NY, USA, 13 pages. <https://doi.org/10.48550/arXiv.2306.08394>

τη βαθιά κατανόηση των νομικών διατάξεων και ηθικών κατευθυντήριων γραμμών. Η εστίαση του δικαίου της ΕΕ στην πρόληψη τόσο των άμεσων όσο και των έμμεσων διακρίσεων απαιτεί μια διαφοροποιημένη προσέγγιση της δικαιοσύνης που υπερβαίνει τα απλά στατιστικά μέτρα. Η εφαρμογή βέλτιστων πρακτικών όπως η δικαιοσύνη από τον σχεδιασμό, οι τακτικοί έλεγχοι, η διαφάνεια και η ανθρώπινη εποπτεία μπορεί να συμβάλει στη γεφύρωση του χάσματος μεταξύ των μετρικών δικαιοσύνης και της νομικής συμμόρφωσης, διασφαλίζοντας ότι τα συστήματα τεχνητής νοημοσύνης λειτουργούν δίκαια και ηθικά εντός του νομικού πλαισίου της ΕΕ⁶⁹.

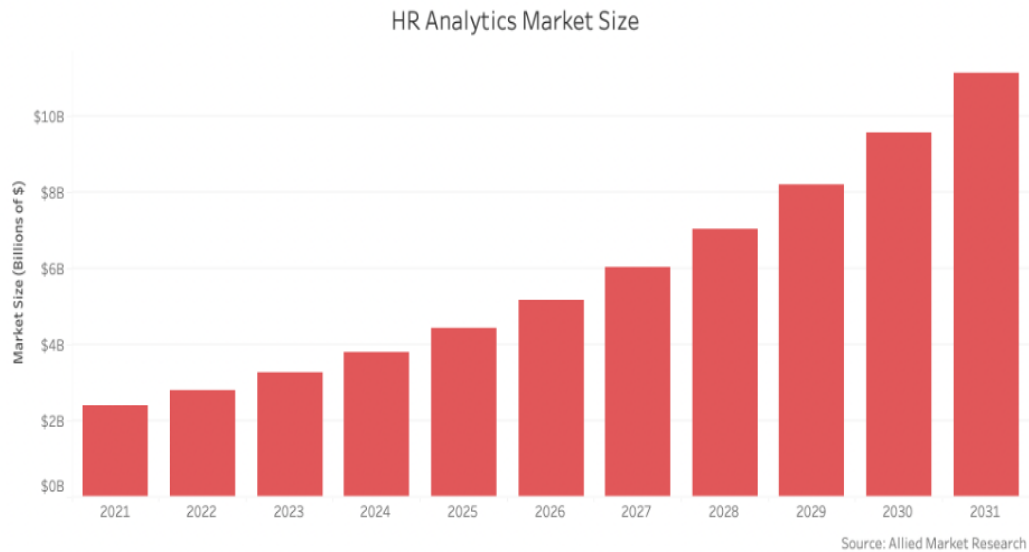
ΚΕΦΑΛΑΙΟ 4– ΕΦΑΡΜΟΓΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΟ ΧΩΡΟ ΤΗΣ ΕΡΓΑΣΙΑΣ ΚΑΙ ΑΛΓΟΡΙΘΜΙΚΗ ΜΕΡΟΛΗΨΙΑ

4.1. Η Αλγοριθμική διοίκηση της εργασίας

Ο όρος «αλγοριθμική διοίκηση της εργασίας» ή αλλιώς «people analytics», «workforce analytics», «HR management by algorithms», «HR Analytics» περιλαμβάνει τις νέες πρακτικές οργάνωσης της εργασίας και διαχείρισης του ανθρώπινου δυναμικού που αξιοποιούν την τεχνητή νοημοσύνη προκειμένου να υποστηρίζεται η διαχείριση, η αξιολόγηση και ο πειθαρχικός έλεγχος των εργαζομένων. Τα αυτοματοποιημένα συστήματα παρακολούθησης και λήψης αποφάσεων που βασίζονται σε αλγόριθμους αντικαθιστούν τα καθήκοντα των διευθυντικών στελεχών όπως η ανάθεση καθηκόντων, η τιμολόγηση των αναθέσεων, ο προσδιορισμός του χρονοδιαγράμματος εργασίας, η παροχή οδηγιών, η αποτίμηση της παρεχόμενης εργασίας, η παροχή κινήτρων και η επιβολή πειθαρχικών μέτρων. Την αλγοριθμική διοίκηση τη συναντάμε κυρίως στη νέα μορφή εργασίας μέσω διαδικτυακών πλατφορμών («sharing/gig/platform economy/crowdfunding»), αλλά πλέον και σε άλλες πιο παραδοσιακές μορφές εργασίας.

⁶⁹ Alvarez, J.M., Colmenarejo, A.B., Elobaid, A. et al., ο.π.

Σύμφωνα με σχετικές προβλέψεις, ο κλάδος HR Analytics θα αναπτυχθεί τα επόμενα δέκα χρόνια, από ένα εκτιμώμενο μέγεθος παγκόσμιας αγοράς 2,4 δισεκατομμυρίων δολαρίων ΗΠΑ το 2021 σε πάνω από 11 δισεκατομμύρια δολάρια ΗΠΑ έως το 2031.



Απεικόνιση του μεγέθους της αγοράς HR Analytics σε δολάρια ΗΠΑ από το 2021 έως το 2031⁷⁰

Σήμερα, οι πάροχοι λογισμικού υπόσχονται στους εργοδότες να αυτοματοποιήσουν τις εξουσίες του εργοδότη καλύπτοντας ολόκληρο τον κύκλο ζωής της εργασιακής σχέσης⁷¹. Ο σκοπός των εφαρμογών αυτών δεν είναι μόνο να αυτοματοποιήσει εργασίες που διαφορετικά θα εκτελούντο από τον άνθρωπο, αλλά και να μεταβιβάσει σε αυτά τα αυτοματοποιημένα συστήματα τη δυνατότητα λήψης, περισσότερο ή λιγότερο, πολύπλοκων, αποφάσεων. Η χρήση αλγορίθμων μπορεί να εξυπηρετήσει διάφορους στόχους που συνδέονται με τη λειτουργία της επιχείρησης όπως την πρόσληψη εργαζομένων, την εκτέλεση της σύμβασης εργασίας και, τέλος, την

⁷⁰ Portillo Chavez K., Bahr J., Vartanian T., (6/12/2022), AI has made its way to the workplace. So how have laws kept pace? <https://oecd.ai/en/work/workplace-regulation-2022> τ.π. 02.05.2024

⁷¹ Adams – Prassl, J., Regulating algorithms at work: Lessons for a “European approach to artificial intelligence”, European Labour Law Journal 2022, Vol. 13 (I) 2022, <https://journals.sagepub.com/doi/full/10.1177/20319525211062558>, τ.π. 27.03.2024.

καταγγελία της⁷². Τα διευθυντικά στελέχη μέσω της ανάλυσης δεδομένων μπορούν να σχεδιάζουν νέες στρατηγικές όσον αφορά τη στόχευση των αγγελιών εργασίας, την πρόσληψη νέων εργαζομένων, τον καθορισμό των αποδοχών, τις προαγωγές, την αξιολόγηση της παραγωγικότητας ακόμα και τις απολύσεις⁷³. Τα νέα ψηφιακά εργαλεία χρησιμοποιούνται για την πρόβλεψη, τη μέτρηση, την καταγραφή και την ανάλυση της αποδοτικότητας των εργαζομένων αλλά και για τη διεξαγωγή πολλών άλλων λειτουργιών στον χώρο εργασίας⁷⁴.

Στα οφέλη της αλγοριθμικής διοίκησης συμπεριλαμβάνονται η ακρίβεια και η αντικειμενικότητα των αλγοριθμικών αποφάσεων, η εξοικονόμηση ανθρωπίνων πόρων και χρόνου και η αποτελεσματικότερη άσκηση των εργοδοτικών εξουσιών. Ωστόσο, οι τεχνολογίες που βασίζονται στην τεχνητή νοημοσύνη επιτρέπουν νέες μορφές παρακολούθησης των εργαζομένων. Στην αλγοριθμική διοίκηση τα προσωπικά δεδομένα των εργαζομένων συλλέγονται από πολυάριθμες συσκευές (π.χ. υπολογιστές εργαζομένων, φορητός εξοπλισμός εργασίας, IoT κ.λ.π.), αναλύονται μέσω αλγορίθμων και αξιοποιούνται για να ληφθούν εν μέρει ή πλήρως αυτοματοποιημένες αποφάσεις⁷⁵. Η χρήση εργαλείων που βασίζονται στα δεδομένα (data – driven tools) οδηγεί στην ενίσχυση και διεύρυνση των εργοδοτικών εξουσιών και των οργανωτικών, ελεγκτικών και πειθαρχικών προνομίων, ενώ περιορίζει σημαντικά την ικανότητα δράσης και ανάληψης πρωτοβουλίας των εργαζομένων⁷⁶, οι οποίοι τελούν υπό στενή επιτήρηση σε ένα πανοπτικό περιβάλλον εργασίας. Η χρήση εργαλείων τεχνητής νοημοσύνης συνεπάγεται τη δεδομενοποίηση (datafication) πτυχών της εργασιακής ζωής, δηλαδή τη μετατροπή τους σε δεδομένα τα οποία χρησιμοποιούνται για να εξαχθούν χρήσιμες γνώσεις ιδιαίτερης αξίας για τις

⁷² Παπαδημητρίου Κ.(2022), Νέες τεχνολογίες, αλγόριθμοι και εργατικό δίκαιο, Εισήγηση στην ΕΣΔΙ, https://www.esdi.gr/wp-content/uploads/2022/seminars/03/dikaio-plir/papadimitriou_2022.pdf, τ.π. 31.08.2024.

⁷³ Aloisi A., ο.π.

⁷⁴ European Parliament, Directorate-General for Internal Policies of the Union, Konle-Seidl, R., Danesi, S., Digitalisation and changes in the world of work – Literature review, European Parliament, 2022, <https://data.europa.eu/doi/10.2861/291260>, τ.π. 27.03.2024.

⁷⁵ Aloisi A., ο.π.

⁷⁶ Aloisi, A., ο.π.

επιχειρήσεις. Οι μέθοδοι που χρησιμοποιούνται συχνά διαταράσσουν το δικαίωμα στην ιδιωτικότητα και την προστασία των προσωπικών δεδομένων των εργαζομένων. Επιπλέον, η ψυχική και σωματική υγεία των εργαζομένων διαταράσσεται καθώς τελούν υπό το άγρυπνο βλέμμα του εργοδότη τους και μετατρέπονται σε μηχανές παραγωγής με μετρήσιμη αποδοτικότητα που σε περίπτωση μη προσδοκώμενης λειτουργίας αποσύρονται.

Επιπρόσθετα, λόγω της αλγοριθμικής προκατάληψης οι αποφάσεις ενός συστήματος τεχνητής νοημοσύνης ενδεχομένως να μην είναι πάντα δίκαιες, για τους υπαλλήλους και τελικά να οδηγούν σε αποτελέσματα που προκαλούν δυσμενείς διακρίσεις, με συνεπακόλουθες δραματικές συνέπειες για τις συνθήκες εργασίας και τα προς το ζην των εργαζομένων. Το πρόβλημα της αλγοριθμικής μεροληψίας διογκώνεται λόγω της δυσκολίας πρόσβασης των εργαζομένων στην πληροφόρηση αλλά και εξαιτίας του ίδιου του τρόπου λειτουργίας ορισμένων αλγορίθμων, ο οποίος είναι αδιαφανής (black box μοντέλα) ώστε είναι σχεδόν αδύνατο για τους εργαζόμενους να κατανοήσουν, πόσο μάλλον να αμφισβητήσουν ανοιχτά τα αποτελέσματά των αλγορίθμων⁷⁷. Οι εργαζόμενοι δεν είναι σε θέση να κατανοήσουν και να διορθώσουν τα αποτελέσματα τέτοιων αποφάσεων όταν δεν γνωρίζουν πώς λειτουργούν τα συστήματα και ποιες μεταβλητές λαμβάνονται υπόψη σε αυτά. Ακόμα και οι ίδιοι οι εργοδότες που προμηθεύονται τα εργαλεία τεχνητής νοημοσύνης που έχουν αναπτύξει διάφορες εταιρείες τεχνολογίας δεν είναι σε θέση να κατανοήσουν τον τρόπο λειτουργίας των αλγορίθμων καθώς πρόκειται για πολύπλοκα μαθηματικά και υπολογιστικά συστήματα. Η έλλειψη διαφάνειας (εγγενής λόγω της λειτουργίας του αλγορίθμου ή ηθελημένη λόγω της απόκρυψης των αναγκαίων πληροφοριών από τους εργοδότες) θέτει σε κίνδυνο τις αρχές της δικαιοσύνης και της λογοδοσίας σε σχέση με την

⁷⁷ Data subjects, digital surveillance, AI and the future of work, European Parliament / Panel for the Future of Science and Technology (STOA) 2020-12 (STOA), [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656305/EPRS_STU\(2020\)656305_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656305/EPRS_STU(2020)656305_EN.pdf), τ.π. 27.03.2024.

απόδοση δικαιωμάτων όπως προαγωγών, άλλων κινήτρων και bonus⁷⁸ αλλά και σε σχέση με όσες εργοδοτικές αποφάσεις επηρεάζουν σημαντικά τους εργαζόμενους. Επιπλέον, συχνά δεν προβλέπεται μηχανισμός επανόρθωσης των εσφαλμένων ή άδικων και μεροληπτικών αλγοριθμικών αποφάσεων, ούτε η δυνατότητα επαναξιολόγησης της απόφασης από άνθρωπο. Μάλιστα, ακόμα και η ίδια η επικοινωνία με άνθρωπο πολλές φορές δεν είναι εξασφαλισμένη.

Ακολούθως, παρατίθενται περιπτώσεις εφαρμογών αλγοριθμικής διοίκησης στον εργασιακό χώρο, κατά τις οποίες παρατηρήθηκαν φαινόμενα αλγοριθμικής μεροληψίας.

4.2. Η εργασία μέσω πλατφόρμας

Οι ψηφιακές πλατφόρμες εργασίας είναι παρούσες σε διάφορους κλάδους της οικονομίας, παρέχοντας υπηρεσίες, όπως η μίσθωση οχημάτων, η παράδοση αγαθών, οι υπηρεσίες καθαρισμού ή φροντίδας και οι υπηρεσίες μετάφρασης.

Οι πλατφόρμες εργασίας παρουσιάζουν πολλά οφέλη καθώς μπορούν να αντιστοιχίσουν αποτελεσματικά την προσφορά και τη ζήτηση εργασίας, να εξοικονομήσουν πρόσθετο εισόδημα μέσω δευτερεύουσας εργασίας σε εργαζόμενους αλλά και να προσφέρουν δυνατότητες βιοπορισμού σε άτομα που αντιμετωπίζουν εμπόδια στην πρόσβαση στην αγορά εργασίας, όπως οι νέοι, τα άτομα με αναπηρία, οι μετανάστες, τα άτομα που προέρχονται από φυλετικές και εθνοτικές μειονότητες ή τα άτομα που αναλαμβάνουν την κατ' οίκον φροντίδα οικείων τους. Επιπλέον, οι πλατφόρμες προσφέρουν καλύτερη οργάνωση της εργασίας, εξοικονόμηση πόρων και χρόνου.

Ωστόσο, καθώς οι ψηφιακές πλατφόρμες εργασίας καθιερώνουν νέες μορφές οργάνωσης της εργασίας, αμφισβητούν τα υφιστάμενα δικαιώματα και υποχρεώσεις

⁷⁸Aloisi, A., ο.π.

που σχετίζονται με το εργατικό δίκαιο και την κοινωνική προστασία⁷⁹. Οι εργαζόμενοι σε πλατφόρμες συχνά εργάζονται ως υπάλληλοι αλλά αντιμετωπίζονται ως ελεύθεροι επαγγελματίες στερούμενοι βασικά εργασιακά δικαιώματα.

Στις ψηφιακές πλατφόρμες εργασίας μίσθωσης οχημάτων, ένας αλγόριθμος και όχι ο εργοδότης αποφασίζει ποια διαδρομή θα αναλάβει ο οδηγός, τι πόσο θα πληρωθεί για τη διαδρομή και αν θα πρέπει να δουλέψει περισσότερο αναθέτοντάς του δρομολόγια. Χαρακτηριστική πρακτική αλγοριθμικής διοίκησης στις πλατφόρμες είναι η «απενεργοποίηση» του προφίλ του εργαζομένου ως πειθαρχική συνέπεια, η οποία συνιστά επί της ουσίας απόλυση. Η «ποινή» αυτή επιβάλλεται κατόπιν αξιολογήσεως του εργαζόμενου μέσω αλγορίθμου, ο οποίος λαμβάνει υπόψη μεταξύ άλλων και τις βαθμολογίες χρηστών της υπηρεσίας⁸⁰. Οι αξιολογήσεις και βαθμολογίες αυτές είναι συνήθως τυποποιημένες και χωρίς αιτιολόγηση, ενώ ο εργαζόμενος δεν έχει τη δυνατότητα να τις αμφισβητήσει. Η ανάγκη διαφάνειας τόσο ως προς τις εργοδοτικές αποφάσεις όσο και τις παραμέτρους που χρησιμοποίησε το σύστημα αξιολόγησης είναι επιβεβλημένη⁸¹. Μάλιστα, όπως έχει διαπιστωθεί τα συστήματα βαθμολόγησης αυτού του τύπου κάθε άλλο παρά δίκαια είναι, καθώς συχνά ενισχύσουν την προκατάληψη των χρηστών αφού ακόμα και αμερόληπτοι χρήστες τελικά επηρεάζονται από τις κακές και μεροληπτικές βαθμολογίες, τις οποίες αντιλαμβάνονται ως δείκτες ποιότητας των υπηρεσιών και ακολουθούν το ίδιο μοτίβο κακής αξιολόγησης⁸². Έτσι, η λήψη αλγοριθμικών αποφάσεων μπορεί, τελικά να βασιστεί σε κοινωνικές προκαταλήψεις και στερεότητες, τα οποία μπορούν να αναπαραχθούν μέσω των αξιολογήσεων των πελατών. Για παράδειγμα έχει

⁷⁹ Πρόταση Οδηγία του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου για τη βελτίωση των συνθηκών εργασίας των εργαζομένων σε πλατφόρμες COM(2021) 762 final, <https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:52021PC0762>, τ.π. 27.03.2024.

⁸⁰ Lukács A., Váradi S, GDPR-compliant AI-based automated decision-making in the world of work, Computer Law & Security Review, Volume 50, September 2023, 105848, <https://www.sciencedirect.com/science/article/pii/S0267364923000584>, τ.π. 27.03.2024.

⁸¹ Παπαδημητρίου, Κ., Γενική Εισαγωγή σε Πληροφορική και Εργατικό Δίκαιο, Νομική Βιβλιοθήκη, 2018, σελ. 5 επ.

⁸² Botelho T. L., Sudhir K., Teng F., Written by Murray S, (14.08.2023), Ratings Systems Amplify Racial Bias on Gig-Economy Platforms, <https://insights.som.yale.edu/insights/ratings-systems-amplify-racial-bias-on-gig-economy-platforms>, τ.π. 27.03.2024.

παρατηρηθεί ότι ορισμένοι χρήστες ταξί αναπαράγουν την προκατάληψη ότι οι γυναίκες δεν είναι καλοί οδηγοί ή ότι οι μετανάστες δεν είναι ευγενικοί αξιολογώντας τους αρνητικά. Οι αξιολογήσεις αυτές από τους πελάτες της πλατφόρμας μπορούν να οδηγήσουν, χωρίς απαραίτητα να οφείλονται σε κακοβουλία, σε ασυνείδητες ή μη ομολογούμενες διακρίσεις που δεν αποκλείεται, στη συνέχεια, να οδηγήσουν στη λήξη της συνεργασίας⁸³. Συνεπώς, ακόμη και οι ουδέτεροι αλγόριθμοι μπορούν να παράγουν μεροληπτικά αποτελέσματα όταν αντικατοπτρίζουν ή ενισχύουν τις υπάρχουσες κοινωνικές προκαταλήψεις.

Αλλά ακόμα και μια καλή βαθμολογία από τους πελάτες δεν αρκεί, όπως φαίνεται, για να εξασφαλίσει κάποιος την εργασία του σε πλατφόρμα. Τον Μάρτιο του 2020 ένας οδηγός της Uber Eats στο Ηνωμένο Βασίλειο, ο οποίος είχε λάβει 96% θετική βαθμολογία από πελάτες, στερήθηκε τελικά την εργασία του στην εταιρεία εξαιτίας ενός «ρατσιστικού» αλγόριθμου. Πιο συγκεκριμένα, ο έγχρωμος οδηγός ανέβασε όπως υποχρεούνταν μια φωτογραφία του (real time selfie) για να επιβεβαιώσει την ταυτότητά του στην εφαρμογή, όμως το λογισμικό απέτυχε να τον αναγνωρίσει, αποκλείοντας την πρόσβασή του στον λογαριασμό του λόγω «ακατάλληλης χρήσης της εφαρμογής Uber». Ειδικότερα, η Uber χρησιμοποιούσε το λογισμικό Microsoft Face API στην εφαρμογή της για να επαληθεύει την ταυτότητα των οδηγών, το οποίο σύμφωνα, με το συνδικάτο ελεύθερων επαγγελματιών Ηνωμένου Βασιλείου («Independent Workers' Union of Great Britain - IWGB») και τους οδηγούς της Uber, δυσκολευόταν να αναγνωρίσει με ακρίβεια άτομα με πιο σκούρες αποχρώσεις δέρματος. Εκπρόσωπος της Microsoft δήλωσε ότι η εταιρεία έχει δεσμευτεί να βελτιώνει το Face API, δίνοντας ιδιαίτερη προσοχή στη δικαιοσύνη και την ακρίβειά του σε σχέση με όλες τις δημογραφικές ομάδες, ενώ ανέφερε ότι παρέχει στους πελάτες της λεπτομερείς οδηγίες και εργαλεία που τους βοηθούν να αμφισβητήσουν την ακρίβεια του συστήματος. Επιπλέον, εκπρόσωπος της Uber ανέφερε ότι το

⁸³ Παπαδημητρίου Κ. (2022), Νέες τεχνολογίες, αλγόριθμοι και εργατικό δίκαιο, Εισήγηση στην ΕΣΔΙ, https://www.esdi.gr/wp-content/uploads/2022/seminars/03/dikaio-plir/papadimitriou_2022.pdf, τ.π. 31.08.2024.

σύστημά παρέχει δυνατότητα ανθρώπινης παρέμβασης για επανεξέταση ώστε να είναι βέβαιο ότι ο αλγόριθμος δεν λαμβάνει αποφάσεις για τα προς το ζην κάποιου αυθαίρετα και χωρίς επίβλεψη⁸⁴. Τελικά, παρά τις διαβεβαιώσεις της εταιρείας, ο οδηγός προσέφυγε δικαστικώς κατά της εταιρείας επικαλούμενος έμμεση διάκριση λόγω φυλής. Ο δικαστικός αγώνας χρηματοδοτήθηκε από την Επιτροπή Ισότητας και Ανθρωπίνων δικαιωμάτων (Equality and Human Rights Commission EHRC) και το Σωματείο εργαζομένων μέσω εφαρμογής οδηγών και ταχυμεταφορέων και ο οδηγός πέτυχε τον εξωδικαστικό συμβιβασμό της υπόθεσής του αλλά και την εκ νέου ενεργοποίηση του λογαριασμού του. Όπως διαφάνηκε ο οδηγός δεν είχε ενημερωθεί για τη διαδικασία που τηρήθηκε για την αναστολή του λογαριασμού του ούτε του είχε παρασχεθεί η δυνατότητα αποτελεσματικής προσφυγής κατά της απόφασης.

Άλλη χαρακτηριστική περίπτωση αλγοριθμικής μεροληψίας σε πλατφόρμα διαπιστώθηκε το 2020 από ιταλικό δικαστήριο⁸⁵, το οποίο έκρινε ότι ο αλγόριθμος της εταιρείας διανομής φαγητού Deliveroo εισήγαγε έμμεση διάκριση, καθώς δεν μπορούσε να αντιμετωπίσει διαφορετικά τις περιπτώσεις που δεν ήταν παρόμοιες κι έτσι τιμωρούσε, μέσω της μείωσης των ανατιθέμενων σε αυτούς παραγγελιών, υπαλλήλους που απουσίαζαν από τη δουλειά τους γιατί απεργούσαν ή ήταν άρρωστοι, περιθωριοποιώντας έτσι τελικά εργαζόμενους που ασκούσαν τα νόμιμα δικαιώματά τους⁸⁶. Είναι γεγονός, ότι οι αλγόριθμοι δεν μπορούν να καταλάβουν αν ένας εργαζόμενος είναι κουρασμένος ή άρρωστος ούτε εξηγούν πάντα γιατί παίρνουν συγκεκριμένες αποφάσεις.

Όπως διαφάνηκε τα εργασιακά δικαιώματα των εργαζομένων στις πλατφόρμες συχνά πλήττονται και οι εργαζόμενοι, οι οποίοι πολλές φορές είναι μετανάστες, υφίστανται δυσμενείς διακρίσεις που οφείλονται στους αλγορίθμους που

⁸⁴Barry, E. (12.10.2021), Uber Drivers say a “Racist” Algorithm is putting them out of work, Time.com, <https://time.com/6104844/uber-facial-recognition-racist/> τ.π. 21.02.2024

⁸⁵Filcams VGIL Bologna and others v. Deliveroo Italia SRL, Court of Bologna, RG 2949/2019, ord. 12.31.2020, η απόφαση είναι διαθέσιμη στον ακόλουθο σύνδεσμο στην ιταλική γλώσσα: <https://www.bollettinoadapt.it/wp-content/uploads/2021/01/Ordinanza-Bologna.pdf>

⁸⁶ Aloisi, A., ο.π.

οργανώνουν την εργασία στις πλατφόρμες. Οι αυτοματοποιημένες αποφάσεις συχνά δεν φέρουν αιτιολόγηση, ενώ απουσιάζει η ανθρώπινη εποπτεία κατά τη διαδικασία λήψης των αποφάσεων και η λογοδοσία. Περαιτέρω, στους εργαζόμενους δεν παρέχεται η δυνατότητα άσκησης αποτελεσματικής προσφυγής.

Για το λόγο αυτό τα θεσμικά όργανα της ΕΕ προκειμένου να βελτιώσουν τις συνθήκες εργασίας των εργαζομένων στις πλατφόρμες εξέδωσαν πρόταση οδηγίας την 09.12.2021⁸⁷, η οποία επιλύει τα ως άνω ζητήματα. Η οδηγία αυτή⁸⁸ υιοθετήθηκε επίσημα σε πρώτη ανάγνωση από το Συμβούλιο της ΕΕ στις 14 Οκτωβρίου 2024 και βρίσκεται στο τελικό στάδιο πριν από την έναρξη ισχύος της. Στην οδηγία αυτή θα κάνουμε αναφορά σε επόμενο κεφάλαιο.

4.3. Αναζήτηση και πρόσληψη εργαζομένων με τη βοήθεια της τεχνητής νοημοσύνης

Με τη βοήθεια της τεχνητής νοημοσύνης ολόκληρη η κρίσιμη διαδικασία της πρόσληψης των εργαζομένων φεύγει από τα χέρια των υπεύθυνων διαχείρισης ανθρώπινου δυναμικού και συχνά ανατίθεται σε πλατφόρμες που λειτουργούν με αλγόριθμους. Μια έρευνα της IBM στα τέλη του 2023 σε περισσότερους από 8.500 παγκόσμιους επαγγελματίες του τομέα της πληροφορικής έδειξε ότι το 42% των εταιρειών χρησιμοποιούν εφαρμογές διαλογής εργαζομένων με τεχνητή νοημοσύνη, ενώ ένα άλλο 40% των ερωτηθέντων εξέταζε θετικά το ενδεχόμενο ενσωμάτωσης τέτοιας τεχνολογίας στη σχετική διαδικασία⁸⁹.

Οι αλγόριθμοι, λοιπόν, αναλαμβάνουν δράση πολύ νωρίς, ήδη, κατά το στάδιο της δημοσίευσης των αγγελιών. Πλατφόρμες όπως το LinkedIn επιτρέπουν στους

⁸⁷ Πρόταση οδηγίας του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου σχετικά με τη βελτίωση των συνθηκών εργασίας στην εργασία σε πλατφόρμες, COM (2021) 762 final της 9^{ης} Δεκεμβρίου 2021.

⁸⁸ European Council, Directive on Improving Working Conditions in Platform Work (PE-89-2024-INIT), Council of the European Union, 2024, <https://data.consilium.europa.eu/doc/document/PE-89-2024-INIT/en/pdf>.

⁸⁹ Τα εργαλεία τεχνητής νοημοσύνης μπορεί να «απορρίπτουν» τους καλύτερους υποψήφιους για θέση εργασίας (16.02.2024), Ναυτεμπορική, <https://www.naftemporiki.gr/society/1592935/ta-ergaleia-technitis-noimosynis-mporei-na-aporriptoyn-toys-kalyteroy-s-ypposifioys-gia-thesi-ergasias/>, τ.π. 27.03.2024.

εργοδότες να προβάλλουν τις αγγελίες τους μόνο σε συγκεκριμένες ομάδες υποψηφίων. Το ποιοι θα μπορούν να δουν τις αγγελίες ορίζεται από τον εργοδότη ή και από έναν αλγόριθμο. Τα κριτήρια επιλογής αυτών που θα μπορούν να δουν στοχευμένα αγγελίες μπορεί να είναι η ηλικία, η εθνικότητα, το φύλο, η εμπειρία ή ακόμα και οι ψηφιακές επαφές με άλλες εταιρείες ή επαγγελματίες. Έτσι, λόγω της αλγοριθμικής μεροληψίας είναι πιθανό να προβάλλονται για παράδειγμα υψηλά αμειβόμενες θέσεις μόνο σε άντρες⁹⁰.

Οι εφαρμογές τεχνητής νοημοσύνης αξιοποιούνται επιπλέον και για τη διαλογή νέων ταλαντούχων υπαλλήλων. Τα λογισμικά που ενσωματώνουν αλγόριθμους μηχανικής μάθησης και πιο συγκεκριμένα προβλεπτικής αναλυτικής μπορούν να ξεχωρίσουν τους κατάλληλους υποψήφιους βάσει βιογραφικών και σχετικών δεικτών απόδοσης, ταιριάζοντας τα χαρακτηριστικά των υποψηφίων όπως η εργασιακή εμπειρία με το ζητούμενο προφίλ εργαζομένου. Παράδειγμα χρήσης τέτοιου λογισμικού συναντάμε στο LinkedIn στην πλατφόρμα του Recruiter⁹¹. Έτσι, η τεχνητή νοημοσύνη βοηθά τους εργοδότες να διαλέξουν ανάμεσα σε εκατοντάδες ή και χιλιάδες αιτήσεις εκείνες τις αιτήσεις με τους πιο κατάλληλους υποψήφιους εξοικονομώντας πολύτιμο χρόνο. Πλέον, η δυνατότητα εξέτασης πολύ μεγάλου αριθμού βιογραφικών καθίσταται δυνατή, προσφέροντας ευκαιρίες σε περισσότερους υποψήφιους, τα βιογραφικά των οποίων κάτω από άλλες συνθήκες ίσως και να μην εξετάζονταν ποτέ, ενώ παράλληλα εξασφαλίζεται και η αντικειμενικότητα. Όμως, τα συστήματα διαλογής υποψηφίων βασίζονται στην μηχανική μάθηση, η οποία αναπαράγει τα μοτίβα που εντοπίζει στα δεδομένα πάνω στα οποία εκπαιδεύτηκε, ενδεχομένως αδικώντας υποψηφίους λόγω αυτών των δεδομένων εκπαίδευσης⁹². Χαρακτηριστικό παράδειγμα είναι η περίπτωση του λογισμικού τεχνητής νοημοσύνης για την διαλογή προσωπικού της Amazon, στην οποία κάναμε αναφορά και πιο πάνω. Το λογισμικό αυτό βαθμολογούσε τα βιογραφικά των υποψηφίων με αστέρια. Ωστόσο η εταιρεία

⁹⁰ De Stefano, V., Wouters, M, ο.π.

⁹¹Ahramovich, A. (12.09.2023), AI in the workplace: 10 key use cases, benefits, and challenges, <https://www.itransition.com/ai/workplace>, τ.π. 27.03.2024.

⁹² De Stefano, V., Wouters, M., ο.π.

διαπίστωσε ότι ο αλγόριθμος μεροληπτούσε σε βάρος των γυναικών τις οποίες απέρριπτε για θέσεις προγραμματιστών. Πιο συγκεκριμένα, το μοντέλο της Amazon είχαν εκπαιδευτεί παρατηρώντας μοτίβα στα βιογραφικά που υποβλήθηκαν στην εταιρεία για μια περίοδο 10 ετών, τα περισσότερα από τα οποία προέρχονταν από άνδρες. Ο αλγόριθμος, λοιπόν, είχε ως εκ τούτου αναπτύξει άμεση αλλά και έμμεση προτίμηση στους άντρες υποψήφιους, απορρίπτοντας βιογραφικά που περιείχαν τη λέξη «γυναίκα» (άμεση προκατάληψη) ή πανεπιστήμια που συνήθως φοιτούν γυναίκες (έμμεση προκατάληψη)⁹³.

Αρκετοί οργανισμοί βασίζονται σε λύσεις μηχανικής μάθησης και πιο συγκεκριμένα προβλεπτικής αναλυτικής και για τη μερική αυτοματοποίηση και υποβοήθηση της διαδικασίας συνεντεύξεων για την επιλογή προσωπικού. Εταιρείες που πωλούν λογισμικό πρόσληψης προσωπικού που χρησιμοποιεί τεχνητή νοημοσύνη ισχυρίζονται ότι μπορούν να διακρίνουν τις ικανότητες και τα χαρακτηριστικά των εργαζομένων υψηλής απόδοσης μέσω συνεντεύξεων με chatbots, βιντεοπαιχνιδιών που καλούνται να παίξουν οι υποψήφιοι και συνεντεύξεων μέσω βιντεοκλήσης όπου αξιολογείται η εκφορά του λόγου, η γλώσσα του σώματος και ο τόνος της φωνής⁹⁴.

Για παράδειγμα, η Tengai⁹⁵ είναι ένας βοηθός τεχνητής νοημοσύνης που αναπτύχθηκε για να κρίνει αντικειμενικά τους υποψηφίους και να διεξάγει διαδικτυακές συνεντεύξεις. Επίσης, η βρετανική πολυεθνική εταιρεία Unilever υιοθέτησε ένα σύστημα προσλήψεων με τεχνητή νοημοσύνη που αναπτύχθηκε από την εταιρεία HireVue, το οποίο μπορεί να αναλύει τη γλώσσα του σώματος των υποψηφίων, τη χρήση λέξεων-κλειδιών και τον τόνο της φωνής των υποψηφίων, μειώνοντας τη διάρκεια της διαδικασίας πρόσληψης από 4 μήνες σε 4 εβδομάδες⁹⁶. Η πλατφόρμα προσλήψεων HireVue παρέχει τη δυνατότητα για ασύγχρονες μέσω βίντεο

⁹³ Κωστή Η., Μπορεί ο αλγόριθμος να είναι δίκαιος; Ο αλγόριθμος και ο «λευκός άνδρας» προκαταλήψεις στην τεχνητή νοημοσύνη, σε συλλογικό έργο Μήτρου Λ., Μπορεί ο αλγόριθμος να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, πανεπιστημιακές εκδόσεις Κρήτης, 2023, σελ.105-106

⁹⁴ De Stefano V, Wouters, M., ο.π.

⁹⁵ Ιστοσελίδα της Tengai: <https://tengai.io/unbiased-recruitment/>

⁹⁶ Ahramovich, A., (12.09.2023), AI in the workplace: 10 key use cases, benefits, and challenges, <https://www.itransition.com/ai/workplace>, τ.π. 27.03.2024.

συνεντεύξεις. Στον υποψήφιο δίνεται μια ερώτηση και 30 λεπτά για να προετοιμάσει την απάντησή του, η οποία βιντεοσκοπείται και αναλύεται από αλγόριθμους τεχνητής νοημοσύνης. Οι αλγόριθμοι οι οποίοι επεξεργάζονται δεδομένα όπως τον τόνο της φωνής και τη γλώσσα του σώματος αποδίδουν ένα σκορ αναφορικά με τα χαρακτηριστικά που απαιτείται να έχει ο υποψήφιος για τη συγκεκριμένη θέση εργασίας (πχ. δημιουργικότητα, οργανωτικότητα, δεξιότητες στην επικοινωνία, ευγένεια κλπ.)⁹⁷.

Το 2020 μια make-up artist της εταιρείας Mac Ηνωμένου Βασιλείου η οποία υποβλήθηκε σε συνέντευξη μέσω της πλατφόρμας HireVue δεν κατάφερε να επαναπροσληφθεί στην εταιρεία όπου εργαζόταν ήδη πριν την περίοδο της πανδημίας covid-19, λόγω αρνητικής βαθμολόγησης της γλώσσας του σώματός της, παρόλο που η γενική απόδοση της είχε κριθεί καλή. Στρεφόμενη δικαστικά κατά της μητρικής εταιρείας της Mac, Estee Lauder, πέτυχε τον συμβιβασμό της υπόθεσής της. Εκπρόσωπος της Mac ισχυρίστηκε ότι στην αξιολόγηση της make-up artist δεν έπαιξε καθοριστικό ρόλο η αναγνώριση προσώπου (αφορούσε το 0,25% της όλης αξιολόγησης) και ότι η εταιρεία διαθέτει ομάδες που αξιολογούν αντικειμενικά τα σχετικά με την απόδοση του εργαζομένου δεδομένα⁹⁸.

Στις συνεντεύξεις αυτές μέσω βίντεο χρησιμοποιούνται εφαρμογές αναγνώρισης προσώπου προκειμένου να εντοπίζονται οι διάφορες εκφράσεις. Μελέτες έδειξαν ότι οι εφαρμογές αναγνώρισης προσώπου ανέλυαν διαφορετικά τη συναισθηματική κατάσταση των ανθρώπων ανάλογα με τη φυλή στην οποία ανήκαν. Η τεχνητή νοημοσύνη συνέδεε τα σκουρόχρωμα πρόσωπα πιο συχνά με αρνητικά συναισθήματα, όπως ο θυμός ή έκρινε τους λευκούς ως περισσότερο ειλικρινείς. Επίσης, όταν οι αλγόριθμοι αυτοί δεν έχουν εκπαιδευτεί με δεδομένα ανθρώπων με

⁹⁷ Kammerer B., *Hired by a Robot: The Legal Implications of Artificial Intelligence Video Interviews and Advocating for greater protection of job applicants*, Volume 107, Iowa Law Review 817 (2022)

⁹⁸ Nachiappan, A., (01.02.2024), *I thought I was going mad!': Former make-up artist says she lost her job after being marked down by AI tool*, <https://news.sky.com/story/amp/companies-increasingly-using-ai-for-recruiting-and-why-it-could-cost-you-your-job-13060905>, τ.π. 27.03.2024.

αναπηρίες, ένας υποψήφιος με αναπηρίες πιθανόν θα αποκλειστεί πχ. λόγω του διαφορετικού τρόπου που κοιτάει, ο οποίος ενδεχομένως θα κριθεί μη ελκρινής. Συνεπώς, τα σχετικά λογισμικά συχνά σφάλουν στην ταυτοποίηση προσώπων αλλά και στην συναγωγή συμπερασμάτων πρόβλεψης συμπεριφορών εξαιτίας της τροφοδότησής τους με χαμηλής ποιότητας δεδομένα και του τρόπου εκπαίδευσης του αλγορίθμου ώστε καταλήγουν σε διακρίσεις σε σχέση με το χρώμα την εθνοτική και φυλετική καταγωγή. Χαρακτηριστικά για την πλατφόρμα Hirevue διαπιστώθηκε ότι απέρριπτε ανθρώπους που δεν είχαν την αγγλική γλώσσα ως μητρική γλώσσα καθώς η πλατφόρμα αδυνατούσε να καταλάβει την προφορά τους⁹⁹.

Τον Ιανουάριο 2021 το Συμβούλιο της Ευρώπης εξέδωσε κατευθυντήριες προς τις κυβερνήσεις, τους νομοθέτες και τις επιχειρήσεις υπό τη Σύμβαση του άρθρου 108 για την προστασία των προσωπικών δεδομένων σύμφωνα με τις οποίες εφαρμογές αναγνώρισης προσώπου οι οποίες συνδέουν την αναγνώριση συναισθημάτων με την πρόσληψη σε εργασία προκαλούν μεγάλη ανησυχία τόσο σε ατομικό όσο και σε κοινωνικό επίπεδο και θα πρέπει να απαγορεύονται¹⁰⁰.

Η Hilke Schellmann ερευνητική δημοσιογράφος και συγγραφέας του βιβλίου “The Algorithm - How AI decides who gets hired, monitored, promoted & fired & why we need to fight back now” αφού δοκίμασε η ίδια εργαλεία πρόσληψης με χρήση τεχνητής νοημοσύνης ανέφερε ότι πολλά από αυτά είναι «μαύρα κουτιά» ως προς τον τρόπο λειτουργίας τους, βασίζονται σε αυθαίρετα και μη επιστημονικά αποδεδειγμένα συμπεράσματα, όπως για παράδειγμα ότι ο τόνος της φωνής μπορεί να φανερώσει πόσο επιτυχημένος θα είναι κάποιος στη δουλειά του, ενώ αλλά καταλήγουν σε διακρίσεις σε βάρος όσων υποβάλλονται στην κρίση τους. Η Schellmann, μάλιστα, καλεί τα τμήματα ανθρώπινου δυναμικού να τηρήσουν σκεπτικιστική στάση απέναντι σε αυτά τα εργαλεία και να τα ελέγχουν. Κατά την άποψή της μια δημόσια εποπτική αρχή θα πρέπει να ελέγχει τα εργαλεία αυτά πριν διατεθούν στην αγορά

⁹⁹ Albaroudi E, Mansouri T, Alameer A. A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. *AI*. 2024; 5(1):383-404. <https://doi.org/10.3390/ai5010019>

¹⁰⁰ Madinier Salis F., A guide to artificial intelligence at the workplace – Your rights on algorithms, European Economic Committee, Report, 2021.

προκειμένου να επιβεβαιώνεται ότι όντως λειτουργούν και ότι δεν καταλήγουν σε διακρίσεις σε βάρος εργαζομένων ή υποψηφίων εργαζομένων. Επίσης, τονίζει ότι οι πωλητές τέτοιων προγραμμάτων θα πρέπει να δημοσιοποιούν τεχνικές εκθέσεις όπου θα εκθέτουν με ποιο τρόπο επικύρωσαν την καλή λειτουργία των εργαλείων αυτών¹⁰¹.

Πλέον, ο Κανονισμός για την τεχνητή νοημοσύνη, στον οποίο θα αναφερθούμε εκτενώς σε επόμενο κεφάλαιο, συμπεριλαμβάνει στις απαγορευμένες χρήσεις, τη χρήση συστημάτων τεχνητής νοημοσύνης για τη συναγωγή συναισθημάτων φυσικού προσώπου στους τομείς του χώρου εργασίας (εκτός εάν η χρήση προορίζεται για ιατρικούς λόγους ή λόγους ασφαλείας) και εντάσσει στα συστήματα υψηλού κινδύνου, αυτά που εφαρμόζονται στην απασχόληση, διαχείριση εργαζομένων και πρόσβαση στην αυτοαπασχόληση, με αποτέλεσμα η χρήση τους να υπόκειται σε αυστηρές προϋποθέσεις και να αντιμετωπίζονται τα ζητήματα που θέτει η Schellmann.

4.4. Πρακτικό παράδειγμα εφαρμογής τεχνητής νοημοσύνης στις εργασιακές σχέσεις

Στο σημείο αυτό και αφού έχουμε ήδη σε προγενέστερα κεφάλαια εξηγήσει πώς λειτουργεί η τεχνητή νοημοσύνη και πώς εισέρχεται στους αλγορίθμους η μεροληψία, θα αναφερθούμε σε μια μελέτη περίπτωσης από τον χώρο των εργασιακών σχέσεων κατά την οποία χρησιμοποιείται μηχανική μάθηση ώστε να διευκολυνθεί η διαδικασία προσλήψεων εργαζομένων¹⁰². Σκοπός μας είναι να αναδείξουμε ποιες

¹⁰¹Corbyn Z (03.02.2024), The AI tools that might stop you getting hired, <https://www.theguardian.com/technology/2024/feb/03/ai-artificial-intelligence-tools-hiring-jobs>, τ.π. 27.03.2024.

¹⁰²Για τη σχετική θεωρία για το πώς λειτουργεί η προβλεπτική αναλυτική βλ. Berry, M. J. A., Linoff, G. S., Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Wiley Editions, 2004. Η ιδέα της μελέτης περίπτωσης για τη διαδικασία υποστήριξης της αυτοματοποιημένης πρόσληψης σε μία εταιρεία βασίστηκε στο άρθρο «Application of Predictive Analytics To Improve The Hiring Process In A Telecommunications Company» των Luh Putu Saraswati, Devia Jayanti και Meditya Wasesa, , Jurnal CoreIT, Vol.8, No.1, June 2022 https://www.researchgate.net/publication/362722810_Application_of_Predictive_Analytics_To_Improve_The_Hiring_Process_In_A_Telecommunications_Company, τ.π. 27.03.2024.

προκλήσεις προκύπτουν προκειμένου να επιτευχθεί μια δίκαιη κρίση με τη βοήθεια των αλγορίθμων.

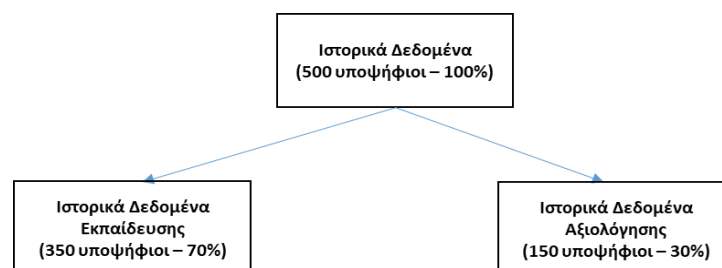
Ας υποθέσουμε λοιπόν ότι μια εταιρεία θέλει να προσλάβει συνεργάτες δικηγόρους για το νομικό της τμήμα. Η διαδικασία πρόσληψης πρέπει να είναι αντικειμενική και να μην στηρίζεται στη διαίσθηση ή υποκειμενική κρίση των αρμοδίων του τμήματος του ανθρώπινου δυναμικού. Για αυτό το λόγο η εταιρεία αποφάσισε να λάβει υποστήριξη χρησιμοποιώντας τεχνικές τεχνητής νοημοσύνης οι οποίες στηρίζονται σε πραγματικά ιστορικά δεδομένα παρελθοντικών προσλήψεων και μεθόδους μηχανικής μάθησης. Το πρώτο στάδιο της διαδικασίας είναι η συλλογή ιστορικών δεδομένων από παρελθοντικές προσλήψεις δικηγόρων που εργάστηκαν στην εταιρεία για τουλάχιστο ένα χρόνο, τηρουμένων σε όλη τη διαδικασία των απαιτήσεων της νομοθεσίας περί προστασίας προσωπικών δεδομένων (άρθρα 5, 6, 15, 22, 25 και 35 ΓΚΠΔ). Τα ιστορικά δεδομένα είναι παρελθοντικά δεδομένα τα οποία αξιοποιούνται για να προβλέψουν μελλοντικά δεδομένα ή τάσεις. Το τμήμα ανθρώπινου δυναμικού σε συνεργασία με το νομικό τμήμα στο οποίο θα εργαστούν οι νέοι δικηγόροι έχει διαπιστώσει ότι ένας εργαζόμενος είναι αποδοτικός και αποτελεσματικός όταν στην ετήσια αξιολόγηση έχει λάβει βαθμό μεγαλύτερο του 80%. Έτσι, λοιπόν, για να προσληφθεί ένας νέος εργαζόμενος θα πρέπει να προβλεφθεί αν στην αξιολόγηση ένα έτος μετά την πρόσληψη θα λάβει βαθμό μεγαλύτερο ή ίσο με το προαναφερθέν κατώφλι. Τα δυο τμήματα συμφώνησαν αξιοποιώντας την εμπειρία τους ότι τα χαρακτηριστικά - μεταβλητές που είναι εν δυνάμει κρίσιμα για να προβλέψουν ότι ένας υποψήφιος θα είναι αποδοτικός (βαθμός $\geq 80\%$) είναι τα έτη προϋπηρεσίας, ο αριθμός ξένων γλωσσών που μιλούν, το όνομα του πανεπιστημίου από το οποίο αποφοίτησαν, ο βαθμός πτυχίου, το επίπεδο γνώσης υπολογιστών, η χρονική διάρκεια που χρειάστηκε για την απόκτηση του πτυχίου, ο αριθμός παραστάσεων σε δικαστήρια, ο αριθμός κερδισμένων δικαστικών υποθέσεων, ο αριθμός εργοδοτών που έχουν αλλάξει και ο τ.κ. του τόπου κατοικίας. Φυσικά σε μια πραγματική περίπτωση

τα χαρακτηριστικά θα ήταν πολύ περισσότερα. Το τμήμα ανθρώπινου δυναμικού, λοιπόν, σε συνεργασία με το τμήμα πληροφορικής συγκέντρωσε ένα αρχείο δεδομένων με τα προαναφερθέντα στοιχεία για 500 εργαζόμενους δικηγόρους. Το αρχείο αυτό περιλαμβάνει τα χαρακτηριστικά των εργαζομένων κατά την ημερομηνία υποβολής του βιογραφικού τους και την απόδοση των εργαζομένων 1 χρόνο μετά την πρόσληψη – μεταβλητή στόχος ($\geq 80\%$ -- > 1 , $\leq 80\%$ -- > 0). Αυτό το αρχείο αποτελεί το ιστορικό σύνολο δεδομένων που θα εκπαιδεύσει το μοντέλο για να εξάγει το αποτέλεσμα, δηλαδή να απαντήσει στο ερώτημα ποια χαρακτηριστικά των εργαζομένων κατά την ημερομηνία υποβολής του βιογραφικού είναι σημαντικά για να προβλέψουν ότι η απόδοση του εργαζομένου ένα χρόνο μετά την πρόσληψη θα είναι μεγαλύτερη του 80% (μεταβλητή στόχος = 1).

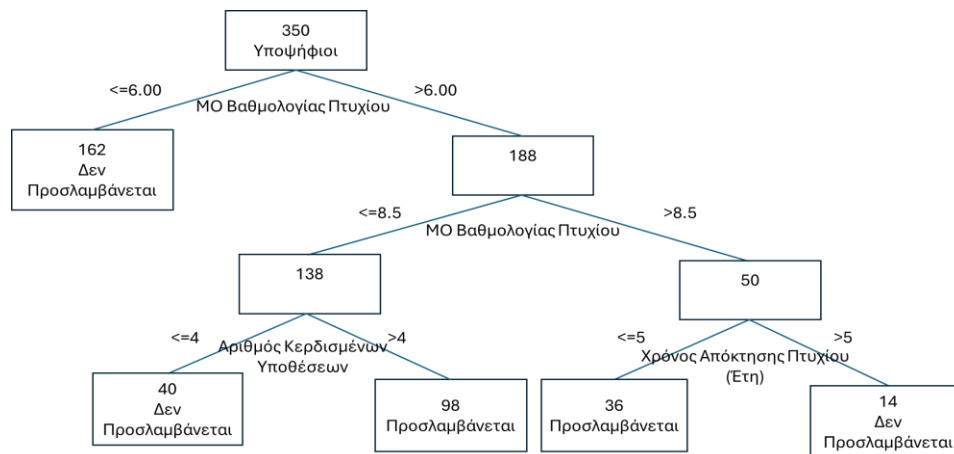
Από το αρχείο των ιστορικών δεδομένων θα εκπαιδευτούν και θα αξιολογηθούν τα μοντέλα πρόβλεψης της τεχνητής νοημοσύνης. Με βάση την αξιολόγησή των μοντέλων θα επιλεγεί το μοντέλο που έχει την καλύτερη απόδοση και το οποίο θα χρησιμοποιηθεί στα μελλοντικά δεδομένα, δηλαδή στα χαρακτηριστικά των μελλοντικών υποψήφιων που θα υποβάλλουν βιογραφικό για να προσληφθούν προκειμένου να προβλέψει ποιοι θα είναι αποδοτικοί στην εργασία τους αν προσληφθούν. Επομένως το μοντέλο θα βοηθήσει στη λήψη της απόφασης της πρόσληψής τους. Μια γενική πρακτική στη μηχανική μάθηση είναι ότι ένας μεγάλος αριθμός αλγορίθμων – μοντέλων εκπαιδεύονται στα ιστορικά δεδομένα και αξιολογούνται από αυτά ως προς την ακρίβειά τους, οπότε στο συγκεκριμένο παράδειγμα θα μπορούσε να χρησιμοποιηθεί ένα δέντρο αποφάσεων, ένα νευρωνικό δίκτυο, μια παλινδρόμηση, ένα τυχαίο δάσος κ.λ.π. Όσο πιο πολλά μοντέλα – αλγόριθμοι εκπαιδευτούν, τόσο αυξάνονται και οι πιθανότητες να βρεθεί ένα μοντέλο το οποίο θα είναι όσο το δυνατό πιο ακριβές στην πρόβλεψή του. Για να γίνει η λεγόμενη αμερόληπτη και ειλικρινή αξιολόγηση των μοντέλων που θα δοκιμαστούν η γενική πρακτική στη μηχανική μάθηση είναι ότι από τα ιστορικά δεδομένα (υποψήφιοι για εργασία) γίνεται τυχαία επιλογή περίπου του 70% ($350 =$

70% X 500) αυτών ώστε με αυτά τα δεδομένα να εκπαιδευτούν τα μοντέλα και επιλέγεται το υπόλοιπο τυχαίο 30% ($150 = 30\% \times 500$) αυτών για να αξιολογηθούν τα μοντέλα όσον αφορά την ακρίβεια τους. Έτσι αν το πρώτο μοντέλο που θα δοκιμαστεί είναι ένα δέντρο αποφάσεων, αυτό θα μάθει από την ανάλυση των 350 υποψήφιων εργαζομένων (ιστορικά δεδομένα εκπαίδευσης) ποια χαρακτηριστικά των υποψήφιων εργαζομένων οδηγούν στο να αποδειχθεί κάποιος υποψήφιος αποδοτικός εργαζόμενος. Η μάθηση αυτή βασίζεται σε στατιστικές μεθοδολογίες. Στη συνέχεια η γνώση που αποκόμισε ο αλγόριθμος από τους 350 εργαζομένους, σε ότι αφορά τη συσχέτιση χαρακτηριστικών των υποψήφιων εργαζομένων κατά τη στιγμή υποβολής βιογραφικού και της απόδοσής τους στην εργασία ένα χρόνο μετά την πρόσληψή τους εφαρμόζεται στους 150 υποψήφιους εργαζόμενους από τα ιστορικά δεδομένα, οι οποίοι θα χρησιμοποιηθούν για αξιολόγηση και παρέχεται η πρόβλεψη για το αν κάθε ένας από αυτούς τους 150 υποψήφιους εργαζόμενους θα αποδειχθεί αποδοτικός ή μη στη θέση εργασίας ένα χρόνο μετά την πρόσληψη. Επειδή όμως για τους 150 αυτούς εργαζόμενους υπάρχουν πραγματικά στοιχεία για το αν ήταν αποδοτικοί ή όχι στην εργασία τους, συγκρίνονται οι προβλέψεις με τα πραγματικά αυτά στοιχεία και προκύπτει έτσι το ποσοστό ακρίβειας του υπό εξέταση μοντέλου (δέντρο απόφασης). Η διαδικασία επαναλαμβάνεται για όλα τα μοντέλα τα οποία θα δοκιμαστούν και επιλέγεται τελικά το μοντέλο το οποίο θα έχει τη μεγαλύτερη ακρίβεια σε ότι αφορά την πρόβλεψη αποδοτικότητας των 150 υποψήφιων εργαζομένων από τα ιστορικά δεδομένα (ιστορικά δεδομένα αξιολόγησης).

Σχήμα στο οποίο φαίνεται ο διαχωρισμός των ιστορικών δεδομένων σε ιστορικά δεδομένα εκπαίδευσης και ιστορικά δεδομένα αξιολόγησης



Υποθετικά, λοιπόν, κατόπιν της δοκιμασίας των διαφόρων μοντέλων το μοντέλο που αποδείχθηκε να έχει την μεγαλύτερη ακρίβεια, έτσι όπως αξιολογήθηκε από τα ιστορικά δεδομένα αξιολόγησης, είναι το παρακάτω δέντρο αποφάσεων:



Το βέλτιστο μοντέλο (δέντρο αποφάσεων)

Το μοντέλο αυτό (δέντρο αποφάσεων) ανήκει στα μοντέλα με την μεγαλύτερη επεξηγησιμότητα, οπότε και είναι δυνατή η ερμηνεία του. Ειδικότερα, μπορεί να εξηγηθεί ποια χαρακτηριστικά θεώρησε το μοντέλο σημαντικά από τα δέκα για να προβλέψει ποιοι υποψήφιοι εργαζόμενοι από αυτούς που θα υποβάλλουν βιογραφικά θα είναι αποδοτικοί και ποιοι όχι, με βάση πιθανολογικά συμπεράσματα. Έτσι, λοιπόν, όπως αποτυπώνεται στο ως άνω σχήμα από τα δέκα χαρακτηριστικά των υποψηφίων εργαζομένων που υπέβαλλαν βιογραφικό το καλύτερο μοντέλο, δηλαδή το δέντρο αποφάσεων, με βάση τη στατιστική ανάλυση των ιστορικών δεδομένων εκπαίδευσης και με βάση την επικύρωσή του από τα ιστορικά δεδομένα αξιολόγησης, κατέληξε στο συμπέρασμα ότι τα πιο σημαντικά χαρακτηριστικά είναι ο μέσος όρος βαθμολογίας πτυχίου, ο αριθμός κερδισμένων δικαστικών υποθέσεων και ο χρόνος

που χρειάστηκε ο υποψήφιος για την αποφοίτηση. Από την ερμηνεία του δέντρου αποφάσεων προκύπτουν οι ακόλουθοι κανόνες:

α) Αν η μέση βαθμολογία πτυχίου του υποψηφίου είναι μικρότερη από 6 (με άριστα το 10) τότε ο δικηγόρος δεν θα πρέπει να προσληφθεί γιατί προβλέπεται ότι δεν θα έχει καλή απόδοση σαν εργαζόμενος.

β) Αν ο υποψήφιος έχει μέση βαθμολογία πτυχίου μεγαλύτερη από 6 (με άριστα το 10) και μικρότερη από 8,5 και αν ο αριθμός των κερδισμένων υποθέσεων που έχει χειριστεί είναι μεγαλύτερος από 4 τότε θα πρέπει να προσληφθεί γιατί προβλέπεται ότι θα έχει καλή απόδοση σαν εργαζόμενος.

γ) Αν ο υποψήφιος έχει μέση βαθμολογία πτυχίου μεγαλύτερη από 6 (με άριστα το 10) και μικρότερη ίση από 8.5 και ο αν ο αριθμός των κερδισμένων υποθέσεων που έχει χειριστεί είναι μικρότερος από 4 τότε δεν θα πρέπει να προσληφθεί γιατί προβλέπεται ότι δεν θα έχει καλή απόδοση σαν εργαζόμενος.

δ) Αν ο υποψήφιος έχει μέση βαθμολογία πτυχίου μεγαλύτερη από 8.5 (με άριστα το 10) και ο χρόνος που χρειάστηκε για να αποφοιτήσει είναι μικρότερος από 5 έτη τότε θα πρέπει να προσληφθεί γιατί προβλέπεται ότι θα έχει καλή απόδοση σαν εργαζόμενος.

ε) Αν ο υποψήφιος έχει μέση βαθμολογία πτυχίου μεγαλύτερη από 8.5 (με άριστα το 10) και αν ο χρόνος που χρειάστηκε για να αποφοιτήσει είναι μεγαλύτερος από 5 έτη δεν θα πρέπει να προσληφθεί γιατί προβλέπεται ότι δεν θα έχει καλή απόδοση σαν εργαζόμενος.

Ακολούθως, 6 μήνες μετά τη δημιουργία του μοντέλου η εταιρεία δημοσιεύει μια αγγελία για την πρόσληψη δικηγόρων και δέχεται 100 αιτήσεις. Οι κανόνες που προέκυψαν από το καλύτερο μοντέλο θα εφαρμοστούν στις 100 αιτήσεις των υποψήφιων εργαζομένων και για κάθε υποψήφιο εργαζόμενο θα εκτιμηθεί από το

μοντέλο η πιθανότητα ένα χρόνο μετά την πρόσληψη του υποψηφίου να αξιολογηθεί ως αποδοτικός.

Στη συνέχεια θα καταταγούν οι υποψήφιοι εργαζόμενοι σε φθίνουσα σειρά ως προς τις πιθανότητες που προέβλεψε το μοντέλο αυτοί να αποδειχθούν αποδοτικοί ένα χρόνο μετά την πρόσληψή τους. Η αποκομισθείσα αυτή γνώση θα βοηθήσει το τμήμα ανθρώπινου δυναμικού να λάβει ενημερωμένες, δίκαιες και τεκμηριωμένες αποφάσεις για τις προσλήψεις. Έτσι, με τη βοήθεια της τεχνολογίας παρέχεται η δυνατότητα στους υπεύθυνους για τις προσλήψεις να μπορούν να διατρέξουν εν τάχει εκατοντάδες βιογραφικά, να εντοπίσουν δεξιότητες και ταλέντα και τελικά να επιλέξουν με αντικειμενικά κριτήρια τους καλύτερους και πιο κατάλληλους υποψήφιους για κάθε θέση εργασίας. Είναι όμως η σύσταση του αλγορίθμου του παραδείγματός μας δίκαιη;

Με δεδομένο ότι το μοντέλο που επιλέχθηκε στο παράδειγμα (δέντρο αποφάσεων) είναι μοντέλο ερμηνεύσιμο μπορούμε να αντιληφθούμε εύκολα ότι ο κανόνας που εξήγαγε ο αλγόριθμος δεν καταλήγει σε δυσμενή διάκριση υποψηφίων λόγω ευαίσθητων χαρακτηριστικών, αφού οι μεταβλητές που επιλέχθηκαν δεν ήταν ευαίσθητες. Ωστόσο, στη μηχανική μάθηση όπως εκθέσαμε γενικά υπάρχει ο κίνδυνος της έμμεσης μεροληψίας μέσω μεταβλητών proxies (πχ. τ.κ.). Συνεπώς, για την εξασφάλιση της δικαιοσύνης στην αλγοριθμική σύσταση θα πρέπει να χρησιμοποιούνται ερμηνεύσιμα μοντέλα. Στην περίπτωση των black box μοντέλων για την ερμηνεία τους υπάρχουν οι στατιστικές μεθοδολογίες LIME, SHAPLEY, τεχνικές μεταμοντελοποίησης, γραφικές μέθοδοι κ.λ.π., οι οποίες για να μπορούν να εφαρμοστούν, όμως, θα πρέπει να υπάρχει πρόσβαση στα ιστορικά δεδομένα εκπαίδευσης και στις προβλέψεις του black box μοντέλου¹⁰³. Περαιτέρω, θα πρέπει να λαμβάνεται ιδιαίτερη πρόνοια ώστε τα δεδομένα εκπαίδευσης να είναι ισορροπημένα και σε περίπτωση που δεν είναι να εξισορροπούνται με στατιστικές τεχνικές. Όταν τα

¹⁰³ Hassija, V., Chamola, V., Mahapatra, A. et al. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cogn Comput* 16, 45–74 (2024). <https://doi.org/10.1007/s12559-023-10179-8>

δεδομένα που χρησιμοποιούνται για την εκπαίδευση ενός μοντέλου είναι ανισοκατανομημένα, δηλαδή όταν έχουμε πολύ περισσότερα παραδείγματα μιας κατηγορίας σε σχέση με μια άλλη, αυτό μπορεί να οδηγήσει σε προκαταλήψεις και λανθασμένες προβλέψεις. Τα δεδομένα θα πρέπει να είναι αντιπροσωπευτικά των διαφορετικών φυλών, φύλων, κοινωνικών υπόβαθρων και πολιτισμών που θα μπορούσαν να επηρεαστούν αρνητικά. Η άνιση κατανομή δεδομένων μπορεί να αντιμετωπιστεί με την επανασταθμιστική δειγματοληψία (oversampling - αύξηση του αριθμού παραδειγμάτων της λιγότερο εκπροσωπούμενης κατηγορίας/undersampling - μείωση του αριθμού παραδειγμάτων της περισσότερο εκπροσωπούμενης κατηγορίας), τεχνική η οποία χρησιμοποιείται στην ανάλυση δεδομένων¹⁰⁴. Κατά τη διαδικασία της μηχανικής μάθησης δεν θα πρέπει να χρησιμοποιούνται ευαίσθητες μεταβλητές ή μεταβλητές που προσεγγίζουν τις ευαίσθητες μεταβλητές (proxies). Στον εντοπισμό των υποκατάστατων μεταβλητών βοηθά μια στατιστική μέθοδος που ονομάζεται ανάλυση συσχέτισης¹⁰⁵, η οποία μπορεί να φανερώσει ποιες μεταβλητές έχουν υψηλή συσχέτιση με ευαίσθητα χαρακτηριστικά. Τέλος, θα πρέπει να χρησιμοποιούνται στατιστικές μετρικές (fairness metrics) για να ελέγχεται αν τα μοντέλα είναι δίκαια απέναντι σε όλες τις ομάδες και αν οι αποφάσεις τους είναι αντικειμενικές. Τέτοιες μετρικές, όπως προαναφέραμε, είναι ενδεικτικά η τεχνική της στατιστικής ισότητας (statistical equality), οι εξισσοροπημένες πιθανότητες – ευκαιρίες (equalized odds) και η δημογραφική ισότητα (demographic parity)¹⁰⁶. Κατά την αξιολόγηση του μοντέλου με τα δεδομένα επικύρωσης και ανάλογα με το τι έχει οριστεί σαν μέτρο δικαιοσύνης (fairness metrics) θα πρέπει να ελέγχεται αν η μέτρηση

¹⁰⁴ Chakraborty, J.; Majumder, S.; Menzies, T. Bias in machine learning software: Why? how? what to do? In Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Athens, Greece, 23–28 August 2021; pp. 429–440 και Albaroudi E, Mansouri T, Alameer A. A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. AI. 2024; 5(1):383-404. <https://doi.org/10.3390/ai5010019>, τ.π. 24.11.2024

¹⁰⁵ Date S (16.08.2022), How to Use Proxy Variables in a Regression Model. Article published in Towards Data Science, <https://towardsdatascience.com/how-to-use-proxy-variables-in-a-regression-model-539f723ab587>

¹⁰⁶ Conor O'Sullivan (05.03.2021), What is Algorithm Fairness? An introduction to the field that aims at understanding and preventing unfairness in machine learning, Towards Data Science, <https://towardsdatascience.com/what-is-algorithm-fairness-3182e161cf9f>, τ.π. 08.08.2024 και Caton S. Haas C. (2023). Fairness in Machine Learning: A Survey. ACM Computing Surveys. 10.1145/3616865, τ.π. 25.08.2024.

δικαιοσύνης (πχ. σφάλμα πρόβλεψης) εξάγει τα ίδια αποτελέσματα σε όλες τις ομάδες που καθορίζονται από τα ευαίσθητα χαρακτηριστικά. Επιπλέον, τα ως άνω μέτρα θα πρέπει να συμπληρώνονται από μέτρα διαφάνειας, ανθρώπινης εποπτείας και λογοδοσίας. Η ανθρώπινη εποπτεία, η διαφάνεια, η λογοδοσία καθώς και η υποχρέωση λήψης μέτρων μετριασμού της αλγοριθμικής μεροληψίας προβλέπονται στη νομοθεσία. Η διαδικασία που ακολουθήθηκε από την εταιρεία στο παράδειγμα μας δεν βρίσκεται εκτός ρυθμιστικού πλαισίου αλλά θα πρέπει να συμμορφώνεται με το νομοθετικό πλαίσιο που θα παραθέσουμε ακολούθως. Έτσι, θα πρέπει να τηρείται από την εργοδότη εταιρεία η εργατική νομοθεσία που απαγορεύει τις διακρίσεις, ο Γενικός Κανονισμός Προστασίας Δεδομένων, ο Κανονισμός για την τεχνητή νοημοσύνη και ο Ν. 4961/2022, νομοθετήματα τα οποία συγκροτούν ένα ισχυρό οπλοστάσιο για την προστασία των εργαζομένων έναντι άδικων αλγοριθμικών αποφάσεων.

ΚΕΦΑΛΑΙΟ 5 – ΤΟ ΝΟΜΙΚΟ ΟΠΛΟΣΤΑΣΙΟ ΓΙΑ ΤΗΝ ΕΞΑΣΦΑΛΙΣΗ ΤΗΣ ΔΙΚΑΙΟΣΥΝΗΣ ΚΑΤΑ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΙΣ ΕΡΓΑΣΙΑΚΕΣ ΣΧΕΣΕΙΣ

5.1. Η εφαρμογή της νομοθεσίας απαγόρευσης των διακρίσεων στη μηχανική μάθηση.

Οι ενσωματωθείσες στην ελληνική έννομη τάξη κοινοτικές Οδηγίες οι οποίες απαγορεύουν τις διακρίσεις¹⁰⁷, είναι διατυπωμένες με τεχνολογικά ουδέτερο τρόπο αφού το πεδίο εφαρμογής τους δεν έχει ως κριτήριο τα μέσα τέλεσης της διακριτικής

¹⁰⁷ Οδηγία 2006/54/ΕΚ του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 5ης Ιουλίου 2006, για την εφαρμογή της αρχής των ίσων ευκαιριών και της ίσης μεταχείρισης ανδρών και γυναικών σε θέματα εργασίας και απασχόλησης (αναδιατύπωση)· Οδηγία 2000/43/ΕΚ του Συμβουλίου, της 29ης Ιουνίου 2000, περί εφαρμογής της αρχής της ίσης μεταχείρισης προσώπων ασχέτως φυλετικής ή εθνοτικής τους καταγωγής· Οδηγία 2000/78/ΕΚ του Συμβουλίου, της 27ης Νοεμβρίου 2000, για τη διαμόρφωση γενικού πλαισίου για την ίση μεταχείριση στην απασχόληση και την εργασία.

μεταχείρισης¹⁰⁸ και συνεπώς μπορούν να εφαρμοστούν και στην περίπτωση που η διάκριση είναι απότοκος ενός αλγορίθμου. Βέβαια, οι αιτίες της αλγοριθμικής μεροληψίας, η αδιαφάνεια στον τρόπο λειτουργίας των συστημάτων τεχνητής νοημοσύνης καθώς και η προστασία των συστημάτων αυτών από την πνευματική ιδιοκτησία και το εμπορικό απόρρητο δημιουργούν νέες προκλήσεις για την εφαρμογή των νομοθεσιών απαγόρευσης των διακρίσεων¹⁰⁹. Η «αυτοματοποιημένη διακριτική μεταχείριση» σε σχέση με τις παραδοσιακές μορφές διακρίσεων είναι πιο αφηρημένη, μη εύληπτη, ανεπαίσθητη και απροσδιόριστη, ώστε είναι δυσχερέστερα διαγνώσιμη και αποδείξιμη¹¹⁰.

Η νομοθεσία απαγόρευσης των διακρίσεων παρέχει μια σειρά διευκολύνσεων στα θύματα διακριτικής μεταχείρισης. Ειδικότερα, προβλέπει τη δικονομική διευκόλυνση της αντιστροφής του βάρους απόδειξης καθώς σε περίπτωση που φέρεται ενώπιον των δικαστικών ή άλλων αρμόδιων αρχών υπόθεση απαγορευμένης διάκρισης, ο θιγόμενος-ενάγων αρκεί να αποδείξει τα πραγματικά περιστατικά από τα οποία τεκμαίρεται η ύπαρξη άμεσης ή έμμεσης διάκρισης λόγω ενός προστατευόμενου γνωρίσματος, ενώ ο εναγόμενος θα πρέπει να αποδείξει ότι η διακριτική μεταχείριση δεν βασίστηκε σε προστατευόμενο χαρακτηριστικό αλλά σε αντικειμενικά κριτήρια. Άλλη δικονομική διευκόλυνση είναι η δυνατότητα εκπροσώπησης του προσώπου που υπέστη τη δυσμενή διάκριση στα δικαστήρια, στις διοικητικές αρχές και όργανα, από οργανώσεις που αποσκοπούν στη διασφάλιση της τήρησης της αρχής της ίσης μεταχείρισης¹¹¹.

Κατά την κρατούσα άποψη, η άμεση διάκριση δεν φαίνεται να αναπτύσσει ιδιαίτερο πεδίο εφαρμογής στις αλγοριθμικές διακρίσεις. Ο λόγος είναι ότι τα ευαίσθητα χαρακτηριστικά κατά το σύνηθες αφαιρούνται από τα δεδομένα εκπαίδευσης αλλά

¹⁰⁸ Παπαγεωργίου Ι., Συστήματα τεχνητής Νοημοσύνης και διακριτική μεταχείριση κατά το προσυμβατικό στάδιο σύναψης της σύμβασης εργασίας, ΕφΑΔΠολΔ, 10/2022, σελ. 1125 – 1140.

¹⁰⁹ Παπαγεωργίου Ι., ο.π.

¹¹⁰ Μήτρου Λ., Μπορεί ο αλγόριθμος να διοικεί; σε συλλογικό έργο Μήτρου Λ. Μπορεί ο αλγόριθμος... να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, Πανεπιστημιακές εκδόσεις Κρήτης, Ηράκλειο Κρήτης, 2023, σελ. 280.

¹¹¹ Κουμεντάκης Σ., ο.π.

και διότι η ίδια η απόδειξη της άμεσης διάκρισης είναι δυσχερής λόγω της δυσκολίας πρόσβασης στην πληροφορία αλλά και των ιδιαιτεροτήτων που παρουσιάζουν τα αλγοριθμικά συστήματα στον τρόπο λειτουργίας τους¹¹².

Ειδικότερα, προϋπόθεση της άμεσης διάκρισης είναι να συνδέεται αιτιωδώς η λιγότερο ευνοϊκή μεταχείριση με ένα προστατευόμενο χαρακτηριστικό. Μια τέτοια περίπτωση θα ίσχυε αν ο υπεύθυνος για τη λήψη της απόφασης τροφοδοτούσε τον αλγόριθμο με ευαίσθητα χαρακτηριστικά και ο αλγόριθμος τα θεωρούσε σημαντικά για την κατηγοριοποίηση¹¹³. Περαιτέρω, σύμφωνα με την αντιστροφή του βάρους απόδειξης στην περίπτωση της άμεσης διάκρισης ο ενάγων θα πρέπει να προσκομίσει ενδείξεις από τις οποίες να μπορεί να συναχθεί – πιθανολογηθεί η αντικειμενική αιτιότητα της διάκρισης, πράγμα το οποίο στην πράξη είναι εξαιρετικά δυσχερές, καθώς ο ενάγων δεν έχει πρόσβαση στο αλγοριθμικό μοντέλο και δεν γνωρίζει τις μεταβλητές που χρησιμοποιήθηκαν. Έτσι, μια γυναίκα υποψήφια που δεν προκρίνεται για συνέντευξη παρά το γεγονός ότι έχει τα κατάλληλα προσόντα εξαιτίας ενός μεροληπτικού, όπως υποψιάζεται, αλγόριθμου αξιολόγησης βιογραφικών βρίσκεται σε δυσχερή θέση να αποδείξει την άμεση σε βάρος της διάκριση καθώς θα πρέπει να αποκτήσει πρόσβαση στον αλγόριθμο της εταιρείας, ώστε να μάθει ποιες μεταβλητές χρησιμοποιήθηκαν για την αξιολόγηση των βιογραφικών. Αλλά ακόμα κι αν αποκτηθεί η πρόσβαση στην πληροφορία με τη βοήθεια του ΓΚΠΔ, στον οποίο θα αναφερθούμε παρακάτω, ο εντοπισμός της αιτιώδους συνάφειας παρουσιάζει ιδιαίτερα προβλήματα λόγω του τρόπου λειτουργίας των αλγορίθμων μηχανικής μάθησης και των τεχνικών εξόρυξης και ανάλυσης δεδομένων, οι οποίοι εντοπίζουν στατιστικές σχέσεις μεταξύ μεταβλητών σε διαθέσιμα δεδομένα, χωρίς να ενδιαφέρονται για το λόγο ύπαρξης αυτών. Ως αποτέλεσμα οι εντοπισθέντες αξιοποιούμενοι συσχετισμοί σπάνια θα αντανακλούν αιτιακές σχέσεις, όπως τις αντιλαμβάνεται ο κοινός νομικός νους, ή συχνά θα

¹¹² Παπαγεωργίου Ι., ο.π.

¹¹³ Παπαγεωργίου Ι., ο.π.

προσκρούουν ως τέτοιες σε φραγμούς της ανθρώπινης λογικής και ηθικής¹¹⁴. Η συσχέτιση στους αλγόριθμους μηχανικής μάθησης δεν σημαίνει πάντα αιτιότητα. Το πρόβλημα απόδειξης της αιτιότητας, δηλαδή η απάντηση στο ερώτημα τι προκαλεί το συγκεκριμένο αποτέλεσμα, δυσχεραίνεται ακόμα περισσότερο στην περίπτωση που χρησιμοποιείται ένα μη διαφανές (black box) μοντέλο μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα στα οποία δεν υπάρχουν σαφείς κανόνες που να εξηγούν γιατί το δίκτυο πήρε μια απόφαση. Ίσως τα δικαστήρια στο μέλλον να κρίνουν ότι η χρήση από τον εργοδότη μη πιστοποιημένων συστημάτων τεχνητής νοημοσύνης, η μη εφαρμογή μέτρων εξάλειψης της αλγοριθμικής μεροληψίας ή η άρνηση πρόσβασης στην αναγκαία πληροφόρηση συνιστούν ενδείξεις για την ύπαρξη διάκρισης λόγω ενός προστατευόμενου χαρακτηριστικού¹¹⁵.

Αντίθετα, η απαγόρευση της έμμεσης διάκρισης παρέχει ένα κατάλληλο προστατευτικό πλαίσιο έναντι των αλγοριθμικών διακρίσεων, καθώς επιτρέπει την ένταξη μεγάλου εύρους περιπτώσεων κατά τις οποίες το αλγοριθμικό σύστημα, παρότι δεν αξιοποιεί άμεσα κάποιο προστατευόμενο χαρακτηριστικό ως παράμετρο για τη λήψη απόφασης, κατ' αποτέλεσμα θίγει σε δυσανάλογο βαθμό μέλη προστατευόμενων ομάδων. Μάλιστα, η όλη διαδικασία της αλγοριθμικής λήψης αποφάσεων μπορεί να θεωρηθεί ως μία ουδέτερη πρακτική κατά την έννοια του ορισμού της έμμεσης διάκρισης, αφού εφαρμόζεται σε όλους αδιακρίτως και στο βαθμό που προστατευόμενα χαρακτηριστικά δεν αξιοποιούνται άμεσα από το αλγοριθμικό σύστημα¹¹⁶. Επιπλέον, η έμμεση διάκριση μετατοπίζει το κέντρο βάρους από την πράξη μεταχείρισης (πχ. εργοδοτική απόφαση) στις δυσμενείς συνέπειες αυτής. Έτσι μετατοπίζεται το κέντρο βάρους από τη λειτουργία και τις παραμέτρους των αλγορίθμων στο αποτέλεσμα αυτών, γεγονός το οποίο διευκολύνει τα θύματα των διακρίσεων ενόψει της αδιαφάνειας αλλά και του τρόπου λειτουργίας των αλγορίθμων. Επιπλέον, στην έμμεση διάκριση δεν απαιτείται η ύπαρξη αιτιώδους

¹¹⁴ Παπαγεωργίου Ι, ο.π.

¹¹⁵ Παπαγεωργίου Ι, ο.π.

¹¹⁶ Παπαγεωργίου Ι., ο.π.

συνάφειας (πχ. φύλο – εργοδοτική απόφαση) όπως στις άμεσες διακρίσεις αλλά μια σχέση μεταξύ ενός φαινομενικά ουδέτερου κριτηρίου (πχ. κριτήριο πλήρους απασχόλησης - proxy) και μιας δυσανάλογης επιβάρυνσης μιας προστατευόμενης ομάδας (βλ. γυναίκες). Έτσι, τα θύματα αποφεύγουν την πρόκληση του «ανοίγματος» του μαύρου κουτιού και μειώνονται οι αποδεικτικές δυσχέρειες. Σημαντικό μέσο για την απόδειξη της έμμεσης διάκρισης από τα θύματα των διακρίσεων είναι η προσκόμιση στατιστικών στοιχείων για την απόδειξη της συγκριτικά δυσμενέστερης μεταχείρισης μιας προστατευόμενης ομάδας. Η αντίδικη πλευρά της εργοδοσίας από την άλλη μπορεί να επικαλεστεί ότι δεν υφίσταται έμμεση διάκριση και η δυσμενέστερη μεταχείριση δικαιολογείται αντικειμενικά καθώς η διάταξη, το κριτήριο ή η πρακτική που συνεπάγεται τη μεταχείριση αυτή υπηρετεί έναν θεμιτό σκοπό και τα μέσα επίτευξής του είναι πρόσφορα και αναγκαία¹¹⁷. Έτσι, στην περίπτωση που μια εταιρεία χρησιμοποιεί αλγόριθμο για να κατατάξει τους εργαζόμενους βάσει της απόδοσής τους, ώστε να κατανείμει μπόνους και ο αλγόριθμος αυτός κατατάσσει συστηματικά τους εργαζόμενους μερικής απασχόλησης (συνήθως γυναίκες) σε χαμηλή βαθμίδα η θιγόμενη εργαζόμενη θα παρουσιάζει στατιστικά δεδομένα που δείχνουν ότι πχ. το 80% των εργαζομένων μερικής απασχόλησης (που είναι κυρίως γυναίκες) κατατάσσονται στο χαμηλότερο 20% της κατάταξης απόδοσης. Αυτό, σε συνδυασμό με την απουσία επαρκούς αιτιολόγησης από την εργοδοσία, μπορεί να οδηγήσει σε αναγνώριση έμμεσης διάκρισης.

5.2. Ο Γενικός Κανονισμός Προστασίας Δεδομένων

Όπως εξηγήσαμε η τεχνητή νοημοσύνη για να λειτουργήσει χρειάζεται μεγάλους όγκους δεδομένων, συμπεριλαμβανομένων και των προσωπικών δεδομένων, τα οποία υποβάλλονται σε αυτοματοποιημένη επεξεργασία οπότε τυγχάνει ομοιόμορφης εφαρμογής σε όλα τα κράτη μέλη της ΕΕ ο Γενικός Κανονισμός Προστασίας Δεδομένων (ΕΕ) 2016/679 (ΓΚΠΔ). Ο ΓΚΠΔ ρυθμίζει το πώς μπορούν οι

¹¹⁷ Παπαγεωργίου Ι., ο.π.

υπεύθυνοι επεξεργασίας, εν προκειμένω οι εργοδότες, να συλλέγουν και να επεξεργάζονται νόμιμα (άρθρα 5, 6 και 9 ΓΚΠΔ) τα προσωπικά δεδομένα των εργαζομένων τους και εξασφαλίζει τη διαφάνεια, τη δικαιοσύνη και τη λογοδοσία στις αυτοματοποιημένες αποφάσεις, αποτελώντας σημαντικό όπλο στα χέρια των εργαζομένων.

Ειδικότερα, σύμφωνα με το άρθρο 5 παρ. 1 περ. α' του ΓΚΠΔ τα δεδομένα προσωπικού χαρακτήρα πρέπει να υποβάλλονται σε νόμιμη θεμιτή - δίκαιη (fair) και διαφανή επεξεργασία. Η δίκαιη επεξεργασία συνεπάγεται την επεξεργασία των προσωπικών δεδομένων κατά τρόπο εύλογα αναμενόμενο από τα υποκείμενα των δεδομένων και όχι με τρόπους που έχουν αδικαιολόγητα αρνητικές επιπτώσεις για αυτά. Συνεπώς, όταν η επεξεργασία προσωπικών δεδομένων με τη χρήση τεχνητής νοημοσύνης οδηγεί σε άδικες διακρίσεις, παραβιάζεται η αρχή της δικαιοσύνης του άρθρου 5 ΓΚΠΔ («αρχή της αντικειμενικότητας»)¹¹⁸. Περαιτέρω, όταν χρησιμοποιούνται μεροληπτικά και ανακριβή ή ελλιπή δεδομένα εκπαίδευσης και το αποτέλεσμα που παράγει ο αλγόριθμος είναι μεροληπτικό παραβιάζεται και μια ακόμα αρχή του ΓΚΠΔ αυτή της ακρίβειας των δεδομένων (άρθρο 5 παρ. 1 περ. δ ΓΚΠΔ) αφού οι βαθμολογίες που συνάγει ο αλγόριθμος έχουν χαμηλή προγνωστική ακρίβεια. Η παραβίαση των αρχών αυτών συνεπάγεται τη δυνητική επιβολή προστίμου στον υπεύθυνο επεξεργασίας από την Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα (ΑΠΔΠΧ) κατ' αρ. 83 ΓΚΠΔ κατόπιν καταγγελίας του εργαζομένου¹¹⁹ ή αυτεπάγγελτης διοικητικής έρευνας της ΑΠΔΠΧ.

Σημαντική προστασία στους εργαζόμενους παρέχει το άρθρο 22 παρ. 1 του ΓΚΠΔ, το οποίο απαγορεύει τις αποφάσεις, που βασίζονται αποκλειστικά σε αυτοματοποιημένη επεξεργασία προσωπικών δεδομένων, συμπεριλαμβανομένης της κατάρτισης προφίλ εφόσον αυτές παράγουν έννομες συνέπειες για το υποκείμενο

¹¹⁸ What about fairness, bias and discrimination?, Ιστοσελίδα Γραφείου Επιτρόπου Πληροφοριών Ηνωμένου Βασιλείου, <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-about-fairness-bias-and-discrimination/>, τ.π. 02.04.2024.

¹¹⁹ Hacker P., ο.π.

των δεδομένων ή το επηρεάζουν σημαντικά. Ο ΓΚΠΔ στο άρθρο 4 περ. 4 κατά τον προσδιορισμό της έννοιας της κατάρτισης προφίλ δίνει έμφαση στην κατάρτιση προφίλ αναφορικά με την απόδοση στην εργασία. Ειδικότερα, προβλέπει ότι κατάρτιση προφίλ συνιστά οποιαδήποτε μορφή αυτοματοποιημένης επεξεργασίας δεδομένων προσωπικού χαρακτήρα που συνίσταται στη χρήση δεδομένων προσωπικού χαρακτήρα για την αξιολόγηση ορισμένων προσωπικών πτυχών ενός φυσικού προσώπου, ιδίως για την ανάλυση ή την πρόβλεψη πτυχών που αφορούν μεταξύ άλλων την απόδοση στην εργασία. Η απαγόρευση του άρθρου 22 ΓΚΠΔ ισχύει μόνο όταν η επεξεργασία παράγει έννομα αποτελέσματα που αφορούν το υποκείμενο των δεδομένων εν προκειμένω, τον εργαζόμενο ή με παρόμοιο τρόπο τον επηρεάζουν σημαντικά (πχ. καταγγελία της σύμβασης εργασίας ή μη πρόσληψη) και μόνο αν η απόφαση λαμβάνεται αποκλειστικά με αυτοματοποιημένο τρόπο χωρίς ανθρώπινη παρέμβαση. Θα πρέπει να αναφέρουμε στο σημείο αυτό ότι προκειμένου μια απόφαση να μη θεωρείται πλήρως αυτοματοποιημένη η ανθρώπινη παρέμβαση θα πρέπει να είναι ουσιαστική και όχι τυπική και να πραγματοποιείται από άτομο που έχει την εξουσία να ανατρέψει την απόφαση¹²⁰ και άρα από άτομο που γνωρίζει και μπορεί να ελέγξει πώς κατέληξε ο αλγόριθμος στο αποτέλεσμα του. Έτσι, οι υπεύθυνοι επεξεργασίας δεν μπορούν να αποφεύγουν την απαγόρευση του άρθρου 22 κατασκευάζοντας μια τυπική ανθρώπινη παρέμβαση¹²¹. Το άρθρο 22, ωστόσο, προβλέπει και εξαιρέσεις στον κανόνα της ως άνω απαγόρευσης. Σύμφωνα, λοιπόν, με την παρ. 2 του άρθρου 22 ΓΚΠΔ κατ' εξαίρεση το υποκείμενο των δεδομένων μπορεί να υπόκειται σε απόφαση που λαμβάνεται αποκλειστικά βάσει αυτοματοποιημένης επεξεργασίας και το επηρεάζει σημαντικά, όταν η απόφαση είναι αναγκαία για τη σύναψη ή την εκτέλεση σύμβασης μεταξύ του υποκειμένου των δεδομένων και του υπευθύνου επεξεργασίας των δεδομένων. Στην εξαίρεση αυτή, θα

¹²⁰ Lukacs A., Varadi S., GDPR – compliant AI – based automated decision – making in the world of work, Computer Law and Security Review 50 (2023) 105848, <https://www.sciencedirect.com/science/article/pii/S0267364923000584>, τ.π. 02.04.2024

¹²¹ Κατευθυντήριες γραμμές για την αυτοματοποιημένη λήψη αποφάσεων και την κατάρτιση προφίλ για τους σκοπούς του Κανονισμού 2016/679 της Ομάδας εργασίας για την προστασία δεδομένων του άρθρου 29, οι οποίες εγκρίθηκαν στις 3 Οκτωβρίου 2017 και αναθεωρήθηκαν και εκδόθηκαν στις 6 Φεβρουαρίου 2018.

μπορούσε να υπαχθεί η σύμβαση εργασίας, εφόσον θα μπορούσε να αιτιολογηθεί η αναγκαιότητα και η αναλογικότητα μιας τέτοιας επεξεργασίας. Όμως, εφόσον συντρέξει αυτή η εξαίρεση η διάταξη προβλέπει ότι πρέπει να προβλέπονται και εγγυήσεις για το υποκείμενο των δεδομένων. Έτσι, ο υπεύθυνος επεξεργασίας, εν προκειμένω ο εργοδότης, θα πρέπει να εφαρμόζει κατάλληλα μέτρα για την προστασία των δικαιωμάτων, των ελευθεριών και των έννομων συμφερόντων του υποκειμένου των δεδομένων, τουλάχιστον του δικαιώματος εξασφάλισης ανθρώπινης παρέμβασης από την πλευρά του υπευθύνου επεξεργασίας, έκφρασης άποψης και αμφισβήτησης της απόφασης. Επιπλέον, οι πλήρως αυτοματοποιημένες αποφάσεις του άρθρου 22 ΓΚΠΔ δεν μπορούν να βασίζονται, πλην ορισμένων εξαιρέσεων που δεν τυγχάνουν εφαρμογής εν προκειμένω, στις ειδικές κατηγορίες δεδομένων προσωπικού χαρακτήρα που αναφέρονται στο αρ. 9 παρ. 1 (βλ. φυλετική ή εθνοτική καταγωγή, πολιτικά φρονήματα, θρησκευτικές ή φιλοσοφικές πεποιθήσεις ή συμμετοχή σε συνδικαλιστική οργάνωση, καθώς και γενετικά δεδομένα και βιομετρικά δεδομένα με σκοπό την αδιαμφισβήτητη ταυτοποίηση προσώπου, δεδομένα υγείας ή που αφορούν τη σεξουαλική ζωή ή τον γενετήσιο προσανατολισμό). Με τον τρόπο αυτό αποτρέπεται ως ένα βαθμό η πιθανότητα μεροληπτικών αλγοριθμικών αποφάσεων.

Το κύριο πρόβλημα στις αποφάσεις που λαμβάνονται με τη χρήση τεχνητής νοημοσύνης είναι η έλλειψη διαφάνειας η οποία έχει ως αποτέλεσμα, εν προκειμένω, οι εργαζόμενοι να δυσκολεύονται να προσφύγουν ενώπιον διοικητικών και δικαστικών αρχών κατά των άδικων αποφάσεων αλλά και γενικότερα να δυσκολεύονται να ασκούν έλεγχο επί των δεδομένων τους και της νομιμότητας επεξεργασίας αυτών. Ο ΓΚΠΔ, όμως, επιβάλλει στους υπεύθυνους επεξεργασίας να τηρούν την αρχή της διαφάνειας (αρ. 5 παρ. 1 περ. α ΓΚΠΔ) και να επεξεργάζονται τα δεδομένα με διαφανή τρόπο σε σχέση με το υποκείμενο των δεδομένων. Ειδικότερα, ο εργοδότης υποχρεούται να παρέχει στους εργαζόμενους, όλες τις σχετικές με την επεξεργασία πληροφορίες σύμφωνα με τα άρθρα 13, 14 και 15 του ΓΚΠΔ και επιπλέον στις περιπτώσεις λαμβάνονται πλήρως αυτοματοποιημένες αποφάσεις θα πρέπει ο

εργοδότης ως υπεύθυνος επεξεργασίας να παρέχει σημαντικές πληροφορίες σχετικά με τη λογική που ακολουθείται καθώς και τη σημασία και τις προβλεπόμενες συνέπειες της εν λόγω επεξεργασίας για το υποκείμενο των δεδομένων. Όπως χαρακτηριστικά αναφέρει η ομάδα εργασίας του άρθρου 29 ο υπεύθυνος επεξεργασίας πρέπει να βρει απλούς τρόπους να εξηγήσει στα υποκείμενα των δεδομένων το σκεπτικό και τα κριτήρια που χρησιμοποιούνται για να ληφθεί η απόφαση. Στο πνεύμα του ΓΚΠΔ, είναι πρωτίστως απαραίτητο να παρέχονται ουσιαστικές πληροφορίες σχετικά με τα παραπάνω και όχι μια περίπλοκη εξήγηση των αλγορίθμων που χρησιμοποιούνται, ο οποίος και δεν χρειάζεται να αποκαλυφθεί πλήρως¹²². Επιπλέον, ο υπεύθυνος επεξεργασίας θα πρέπει να παρέχει στοιχεία επικοινωνίας ώστε το υποκείμενο των δεδομένων να μπορεί να ζητήσει επανεξέταση της απόφασης σύμφωνα με τις διατάξεις του άρθρου 22 παράγραφος 3¹²³.

Ο υπεύθυνος επεξεργασίας έχει επιπλέον και το καθήκον της λογοδοσίας (αρ. 5 παρ. 2 ΓΚΠΔ) και θα πρέπει να είναι σε θέση να αποδείξει ότι τηρεί τις αρχές του άρθρου 5 ΓΚΠΔ. Ένα σημαντικό εργαλείο λογοδοσίας είναι η κατ' αρ. 35 ΓΚΠΔ μελέτη εκτίμησης αντικτύπου σχετικά με την προστασία δεδομένων (ΕΑΠΔ) η οποία πρέπει να εκπονείται από τον υπεύθυνο επεξεργασίας όταν αυτός σκοπεύει να προβεί σε επεξεργασία η οποία συνεπάγεται υψηλό κίνδυνο για τα δικαιώματα των υποκειμένων των δεδομένων. Εν προκειμένω, η χρήση αλγορίθμων τεχνητής νοημοσύνης συνεπάγεται υψηλό κίνδυνο για τα δικαιώματα των υποκειμένων των δεδομένων, στους κινδύνους δε αυτούς περιλαμβάνεται και ο κίνδυνος διακρίσεων. Συνεπώς ο εργοδότης ως υπεύθυνος επεξεργασίας θα πρέπει να εκπονεί σύμφωνα με το αρ. 35 ΓΚΠΔ μελέτη εκτίμησης αντικτύπου σχετικά με την προστασία δεδομένων (ΕΑΠΔ) κατά την οποία θα πρέπει μάλιστα να ζητείται και η γνώμη των

¹²² Κατευθυντήριες γραμμές για την αυτοματοποιημένη λήψη αποφάσεων και την κατάρτιση προφίλ για τους σκοπούς του Κανονισμού 2016/679 της Ομάδας εργασίας για την προστασία δεδομένων του άρθρου 29, οι οποίες εγκρίθηκαν στις 3 Οκτωβρίου 2017 και αναθεωρήθηκαν και εκδόθηκαν στις 6 Φεβρουαρίου 2018.

¹²³ Lukacs A., Varadi S., GDPR – compliant AI – based automated decision – making in the world of work, Computer Law and Security Review 50 (2023) 105848, <https://www.sciencedirect.com/science/article/pii/S0267364923000584>, τ.π. 02.04.2024

εργαζομένων και των εκπροσώπων των εργαζομένων ώστε να υπάρχει αυξημένη διαφάνεια¹²⁴. Η ΕΑΠΔ αξιολογεί του κινδύνους που προκύπτουν από την επεξεργασία για τα δικαιώματα και τις ελευθερίες των φυσικών προσώπων και προτείνει μέτρα για τον περιορισμό τους. Συνεπώς η Εποπτική Αρχή (ΑΠΔΠΧ) θα πρέπει, στο πλαίσιο ελέγχου μιας ΕΑΠΔ, να διασφαλίζει ότι λαμβάνονται επαρκή μέτρα αλγοριθμικής δικαιοσύνης για τον εντοπισμό και την εξάλειψη της αλγοριθμικής μεροληψίας ήδη στο στάδιο εκπαίδευσης του αλγορίθμου¹²⁵.

Άλλωστε η προστασίας των δεδομένων από τον σχεδιασμό και εξ ορισμού κατ' αρ. 25 ΓΚΠΔ συνεπάγεται την εφαρμογή κατάλληλων τεχνικών και οργανωτικών μέτρων αφού ληφθούν υπόψη οι κίνδυνοι για τα δικαιώματα και τις ελευθερίες των υποκειμένων των δεδομένων προκειμένου να εφαρμόζονται οι αρχές προστασίας των δεδομένων¹²⁶. Μάλιστα, στην αιτιολογική σκέψη 71 ΓΚΠΔ αναφέρεται ότι ο υπεύθυνος επεξεργασίας για να διασφαλιστεί η δίκαιη και διαφανής επεξεργασία των δεδομένων θα πρέπει να χρησιμοποιεί κατάλληλες μαθηματικές ή στατιστικές διαδικασίες για την κατάρτιση προφίλ και να λαμβάνει τεχνικά και οργανωτικά μέτρα ώστε οι παράγοντες που οδηγούν σε ανακρίβειες στα προσωπικά δεδομένα να διορθώνονται, ο κίνδυνος σφαλμάτων να ελαχιστοποιείται και να καθίστανται ασφαλή τα προσωπικά δεδομένα ώστε να αποτρέπονται οι διακρίσεις σε βάρος φυσικών προσώπων.

Η Εποπτική Αρχή προστασίας προσωπικών δεδομένων του Ηνωμένου Βασιλείου (ICO) συμβουλεύει σχετικά τις επιχειρήσεις – υπεύθυνους επεξεργασίας, εφόσον εμπίπτουν σε μια από τις προβλεπόμενες εξαιρέσεις όπου επιτρέπονται τις πλήρως αυτοματοποιημένες αποφάσεις, να παρέχουν στα υποκείμενα των δεδομένων

¹²⁴Lukacs A., Varadi S., GDPR – compliant AI – based automated decision – making in the world of work, Computer Law and Security Review 50 (2023) 105848, <https://www.sciencedirect.com/science/article/pii/S0267364923000584>, τ.π. 02.04.2024

¹²⁵ Hacker P., ο.π.

¹²⁶ What about fairness, bias and discrimination?, Ιστοσελίδα Γραφείου Επιτρόπου Πληροφοριών Ηνωμένου Βασιλείου, <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-about-fairness-bias-and-discrimination/>, τ.π. 02.04.2024.

πληροφορίες σχετικά με την επεξεργασία αλλά και την δυνατότητα να ζητούν ευχερώς ανθρώπινη παρέμβαση ή να αμφισβητήσουν την απόφαση και συστήνει να προβαίνουν οι υπεύθυνοι επεξεργασίας σε τακτικούς ελέγχους για να επιβεβαιώνουν ότι τα συστήματά τους λειτουργούν όπως προβλέπεται¹²⁷. Η ίδια Εποπτική Αρχή στις 12 Δεκεμβρίου 2023, δημοσίευσε προσχέδιο κατευθυντήριων οδηγιών προκειμένου να βοηθήσει τους εργοδότες να είναι σε συμμόρφωση με τη νομοθεσία προστασίας προσωπικών δεδομένων κατά τη διαδικασία πρόσληψης και επιλογής προσωπικού, περιλαμβάνοντας και την περίπτωση της χρήσης εργαλείων τεχνητής νοημοσύνης. Σύμφωνα, λοιπόν, με τις οδηγίες αυτές οι εργοδότες δεν θα πρέπει να κάνουν διαλογή και αξιολόγηση των υποψηφίων μέσω αποκλειστικά αυτοματοποιημένης λήψης αποφάσεων και δημιουργίας προφίλ, εκτός εάν μπορούν να βασίζονται σε κάποια εξαίρεση του άρθρου 22 UK GDPR (αναγκαιότητα σύναψης σύμβασης με τον υποψήφιο, ή ρητή συγκατάθεση ή νομοθετική πρόβλεψη) και θα πρέπει να διαθέτουν δικλείδες ασφαλείας όπως η παροχή στα υποκείμενα των δεδομένων του δικαιώματος να ζητήσουν ανθρώπινη παρέμβαση, να εκφράσουν την άποψή τους και να αμφισβητήσουν την απόφαση. Στην περίπτωση της αποκλειστικά αυτοματοποιημένης λήψης αποφάσεως και δημιουργίας προφίλ, εφόσον συντρέχουν οι ως άνω προϋποθέσεις, οι εργοδότες πρέπει να τηρούν τις υποχρεώσεις διαφάνειας παρέχοντας ουσιαστικές πληροφορίες στους υποψηφίους σχετικά με τη λογική και τη σημασία της απόφασης και τις πιθανές συνέπειες για αυτούς. Όταν λαμβάνουν υπόψη οι εργοδότες εν μέρει αυτοματοποιημένες αποφάσεις και την δημιουργία προφίλ, πρέπει να υπάρχει ουσιαστική ανθρώπινη συμμετοχή στη διαδικασία από πρόσωπο το οποίο έχει την εξουσία να διαφοροποιηθεί και να ανατρέψει τις συστάσεις της τεχνητής νοημοσύνης. Σε περίπτωση που οι εργοδότες χρησιμοποιούν πάροχο υπηρεσιών τεχνητής νοημοσύνης τονίζεται από την Εποπτική Αρχή ότι θα πρέπει να ορίζεται ο ρόλος του ανά επεξεργασία (υπεύθυνος επεξεργασίας, από

¹²⁷ Rights related to automated decision making including profiling, Ιστοσελίδα Γραφείου Επιτρόπου Πληροφοριών Ηνωμένου Βασιλείου (ICO), <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/rights-related-to-automated-decision-making-including-profiling/>, τ.π. 02.04.2024

κοινού υπεύθυνος επεξεργασίας ή εκτελών την επεξεργασία). Τέλος, αναφέρεται ότι όταν η τεχνητή νοημοσύνη χρησιμοποιείται για τις προσλήψεις, πρέπει να διενεργείται ΕΑΠΔ καθώς μια τέτοια επεξεργασία είναι πιθανό να οδηγήσει σε υψηλό κίνδυνο για τα δικαιώματα και τις ελευθερίες των υποψηφίων. Μέσω των οδηγιών αυτών η Εποπτική Αρχή βοηθά τους εργοδότες όχι μόνο στο να είναι σε συμμόρφωση με τη νομοθεσία των προσωπικών δεδομένων αλλά και να λάβουν υπόψη τις ηθικές συνέπειες της χρήσης της τεχνητής νοημοσύνης κατά τις προσλήψεις¹²⁸.

Εάν, επομένως, η αλγοριθμική απόφαση παραβιάζει την αρχή της δικαιοσύνης, της διαφάνειας και της ακρίβειας (αρ. 5 ΓΚΠΔ) και το αρ. 22 παρ. 3 ΓΚΠΔ περί λήψης μέτρων, οι Εποπτικές Αρχές, διαθέτουν σειρά μέτρων για την επιβολή της αλγοριθμικής δικαιοσύνης. Οι Εποπτικές Αρχές έχουν την εξουσία να προβούν σε ελέγχους κατ' αρ. 58 παρ. 1 β ΓΚΠΔ. Η ομάδα εργασίας του άρθρου 29 ορίζει ότι στους ελέγχους περιλαμβάνονται δοκιμές των αλγορίθμων που χρησιμοποιούνται και αναπτύχθηκαν από συστήματα μηχανικής μάθησης για να επιβεβαιωθεί ότι πράγματι αποδίδουν τα προσδοκώμενα και δεν παράγουν μεροληπτικά, λανθασμένα ή αδικαιολόγητα αποτελέσματα. Η εποπτική αρχή θα μπορούσε, μάλιστα, να χρησιμοποιήσει τις διορθωτικές της εξουσίες σύμφωνα με το άρθρο 58 παράγραφος 2 ΓΚΠΔ για την επιβολή στρατηγικών ελαχιστοποίησης της μεροληψίας¹²⁹.

Η χρησιμότητα του ΓΚΠΔ για την προστασία των εργαζομένων από άδικες αποφάσεις που λήφθηκαν στο πλαίσιο της αλγοριθμική διοίκηση των εργασιακών σχέσεων διαφάνηκε στις ακόλουθες δικαστικές υποθέσεις. Ειδικότερα, τον Μάρτιο του 2021, το Περιφερειακό Δικαστήριο του Άμστερνταμ έκρινε ότι η πρακτική του συστήματος της Ola μιας πλατφόρμας μίσθωσης ταξί που επέβαλε χρηματικές κυρώσεις σε όσους οδηγούς ακύρωναν δρομολόγια συνιστούσε αυτοματοποιημένη απόφαση κατά την έννοια του άρθρου 22 ΓΚΠΔ και για αυτό και η εταιρεία έπρεπε να

¹²⁸Kourti P., ICO Issues Draft Guidance on Employment Records and the Recruitment and Selection Process, 20 February 2024, <https://trilateralresearch.com/uncategorized/ico-issues-draft-guidance-on-employment-records-and-the-recruitment-and-selection-process>, τ.π. 02.04.2024.

¹²⁹Hacker P., ο.π.

παρέχει στους οδηγούς αιτιολόγηση της απόφασης και πληροφορίες για τα δεδομένα και κριτήρια που λήφθηκαν υπόψη από τον αλγόριθμο¹³⁰. Τον Ιούνιο του 2021, η ιταλική εποπτική αρχή, επέβαλε πρόστιμο ύψους άνω των 2,6 εκατομμυρίων ευρώ στην ψηφιακή πλατφόρμα Foodinho για τη διακριτική μεταχείριση των οδηγών της μέσω αλγορίθμου. Μεταξύ των παραβάσεων που της καταλογίστηκαν ήταν ότι δεν κατάφερε να παράσχει διαφανείς πληροφορίες σχετικά με το σύστημα αξιολόγησης της φήμης των οδηγών της και ότι δεν παρείχε στους οδηγούς της το δικαίωμα να αμφισβητήσουν την αλγοριθμική λήψη αποφάσεων ή να ζητήσουν ανθρώπινη παρέμβαση σε αποφάσεις που επηρέασαν σημαντικά τις ευκαιρίες τους για απασχόληση¹³¹. Όπως μάλιστα αποκάλυψε η σχετική έρευνα το σύστημα της Foodinho κατέληγε σε διακρίσεις που απέκλειαν τους οδηγούς από ευκαιρίες για δουλειά¹³². Τον Ιούλιο 2021 η ιταλική εποπτική επέβαλε πρόστιμο 2,5 εκατομμυρίων ευρώ στην ψηφιακή πλατφόρμα Deliveroo διότι δεν ήταν διαφανής απέναντι στους οδηγούς σχετικά με το σύστημα που χρησιμοποιούσε για την ανάθεση βαρδιών και διότι συνέλλεγε μη αναγκαία προσωπικά δεδομένα των οδηγών παραβιάζοντας την αρχή της νομιμότητας, διαφάνειας και ελαχιστοποίησης¹³³.

Συνεπώς, ο ΓΚΠΑ αναθέτει το βάρος στον εργοδότη να αναπτύξει διαδικασίες που δεν είναι μόνο αποτελεσματικές από οργανωτική άποψη αλλά και διαφανείς και δίκαιες, ενώ παρέχει στους εργαζόμενους ένα σύνολο δικαιωμάτων που μπορούν να χρησιμοποιηθούν τόσο για την αμφισβήτηση και την αλλαγή της τελικής απόφασης

¹³⁰The Ola & Uber judgments: for the first time a court recognises a GDPR right to an explanation for algorithmic decision-making, (28.04.2021), <https://eulawanalysis.blogspot.com/2021/04/the-ola-uber-judgments-for-first-time.html>, τ.π. 28.08.2024 και Lazcoz, G. (2021). Automated Decision-Making Under Amsterdam's District Court Judgements : Drivers v. Uber and Ola, https://www.researchgate.net/publication/356531942_Automated_Decision-Making_Under_Amsterdam's_District_Court_Judgements_Drivers_v_Uber_and_Ola, τ.π. 29.08.2024

¹³¹ Portillo Chavez K., Bahr J., Vartanian T., (06.12.2022), AI has made its way to the workplace. So how have laws kept pace? <https://oecd.ai/en/wonk/workplace-regulation-2022> τ.π. 02/05/2024.

¹³² Salvi del Pero, A., P. Wyckoff and A. Vourc'h (2022), "Using Artificial Intelligence in the workplace: What are the main ethical risks?", OECD Social, Employment and Migration Working Papers, No. 273, OECD Publishing, Paris, <https://doi.org/10.1787/840a2d9f-en>, τ.π. 21.04.2024.

¹³³ Garante per la protezione dei dati personali Italy – national case no 9685994/22-07-2021.

όσο και για την προετοιμασία του εδάφους για μια αγωγή που βασίζεται στη νομοθεσία που απαγορεύει τις διακρίσεις¹³⁴.

5.3. Ο Κανονισμός για την τεχνητή νοημοσύνη.

Στις 13.03.2024 ψηφίστηκε από το Ευρωκοινοβούλιο ο Ευρωπαϊκός Κανονισμός για την Τεχνητή Νοημοσύνη¹³⁵, ο οποίος θεσπίζει ένα ενιαίο νομικό πλαίσιο για την ανάπτυξη, εμπορία και χρήση της τεχνητής νοημοσύνης σύμφωνα με τις αξίες της Ευρωπαϊκής Ένωσης. Πρόκειται για τον πρώτο ολοκληρωμένο νόμο παγκοσμίως που ρυθμίζει την τεχνητή νοημοσύνη¹³⁶. Στο μέλλον είναι βέβαιο ότι θα θεσπιστούν αντίστοιχες νομοθεσίες και από τρίτες χώρες εφόσον αυτές επιθυμούν να συνεχίσουν να δραστηριοποιούνται στην ευρωπαϊκή αγορά. Το φαινόμενο αυτό παρατηρήθηκε και με τον ΓΚΠΔ και είναι γνωστό ως το φαινόμενο της επίδρασης των Βρυξελλών («Brussels effect»). Όπως τονίζει η καθηγήτρια Λίλιαν Μήτρου σημαντικό στοιχείο του Κανονισμού είναι το μέτρο της εξωεδαφικής εφαρμογής του. Ειδικότερα, ο Κανονισμός θα εφαρμοστεί στους παρόχους, οι οποίοι θα διαθέτουν συστήματα τεχνητής νοημοσύνης σε χώρα της ΕΕ ακόμα κι αν έχουν την εγκατάστασή τους εκτός ΕΕ, θα εφαρμοστεί σε χρήστες συστημάτων εντός της ΕΕ, ανεξάρτητα από το πού έχει παραχθεί το σύστημα, αλλά ακόμα και σε παρόχους ή χρήστες συστημάτων τεχνητής νοημοσύνης που βρίσκονται σε τρίτη χώρα, όταν τα στοιχεία εξόδου που παράγει το σύστημα (τα αποτελέσματα της ανάλυσης δηλαδή) χρησιμοποιούνται στην ΕΕ¹³⁷. Ο Κανονισμός μετά την ψήφισή του δημοσιεύθηκε στην Επίσημη Εφημερίδα της Ευρωπαϊκής Ένωσης στις 12 Ιουλίου 2024, θα αρχίσει να εφαρμόζεται 2

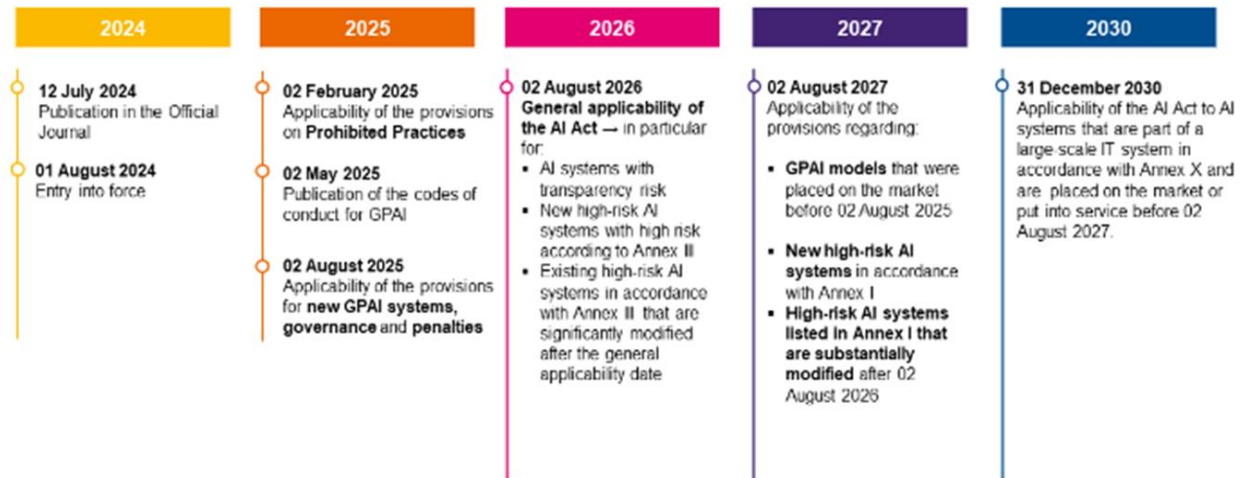
¹³⁴ Aloisi A, ο.π.

¹³⁵ P9_TA(2024)0138 ο.π.

¹³⁶ Η κυβέρνηση Μπάιντεν στις ΗΠΑ εξέδωσε τον Οκτώβριο 2023 ένα εκτελεστικό διάταγμα για τη διαχείριση των κινδύνων που απορρέουν από την τεχνητή νοημοσύνη βλ. Fact Sheet 30.10.2023 President Biden issues executive order on safe, secure, and trustworthy Artificial Intelligence, whitehouse.gov, <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>, τ.π. 02.04.2024.

¹³⁷ Τεχνητή Νοημοσύνη: Οι αλλαγές στη νομοθεσία - Έξι ειδικοί εξηγούν (17.03.2024), <https://www.sofokleousin.gr/texniti-noimosyni-oi-allages-sti-nomothesia-eksi-eidikoi-eksigoun>, τ.π. 02/04/2024.

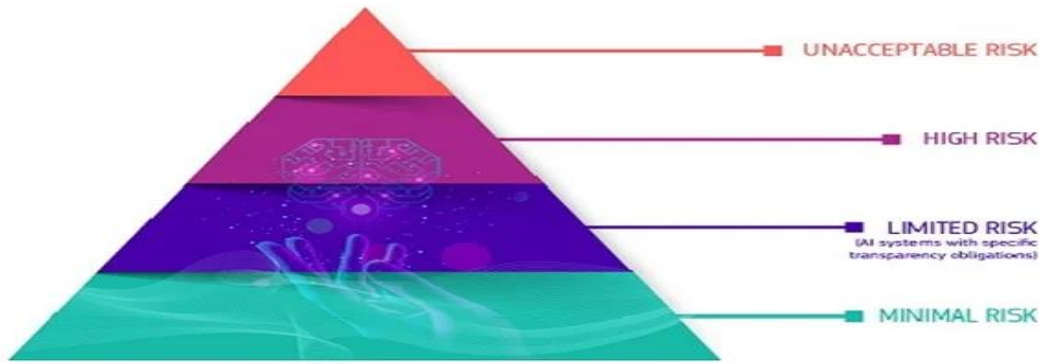
χρόνια μετά την έναρξη ισχύος του (2 Αυγούστου 2026) και περιλαμβάνει ένα περίπλοκο σύστημα προθεσμιών και μεταβατικών διατάξεων.



Απεικόνιση με απλοποιημένους όρους του συστήματος προθεσμιών του Κανονισμού για την Τεχνητή νοημοσύνη¹³⁸.

Στην Ελλάδα σύμφωνα με το από 12.11.2024 δελτίο τύπου του Υπουργείου Ψηφιακής Διακυβέρνησης αρμόδιες αρχές προστασίας των θεμελιωδών δικαιωμάτων σε σχέση με τη χρήση της τεχνητής νοημοσύνης στην Ελλάδα θα είναι η ΑΠΔΠΧ, ο ΣτΠ, η ΑΔΑΕ και η ΕΕΔΑ. Η Ευρωπαϊκή Επιτροπή λαμβάνοντας υπόψη τα οφέλη αλλά και τους κινδύνους της τεχνητής νοημοσύνης επέλεξε μια προσέγγιση με βάση τον κίνδυνο (risk – based approach) που προκύπτει από τα συστήματα τεχνητής νοημοσύνης για την κοινωνία και τους ανθρώπους, καταλήγοντας σε τέσσερις διαβαθμίσεις κινδύνου (μη αποδεκτός, υψηλός, περιορισμένος και χαμηλός κίνδυνος).

¹³⁸ EU’s AI Act published in Official Journal, transition periods now known, 12/07/2024, <https://cms-lawnow.com/en/ealerts/2024/07/eu-s-ai-act-published-in-official-journal-transition-periods-now-know>, τ.π. 12/07/2024.



Απεικόνιση των τεσσάρων επιπέδων κινδύνου στα συστήματα τεχνητής νοημοσύνης¹³⁹

Στην κορυφή της πυραμίδας, όπως αποτυπώνεται στο ως άνω σχήμα, βρίσκονται τα συστήματα μη αποδεκτού κινδύνου, τα οποία θεωρούνται απειλή για την ασφάλεια και τα θεμελιώδη δικαιώματα των ανθρώπων και για αυτό απαγορεύεται να διατίθενται στην ΕΕ. Στα συστήματα αυτά περιλαμβάνονται μεταξύ άλλων συστήματα κοινωνικής βαθμολόγησης, όπως αυτά που χρησιμοποιεί η Κίνα, συστήματα που χρησιμοποιούν υποσυνείδητες πρακτικές χειραγώγησης ή εκμεταλλεύονται ευάλωτες κοινωνικές ομάδες αλλά και συστήματα αναγνώρισης συναισθημάτων στον χώρο εργασίας (πχ. συστήματα όπως αυτά που αναφέραμε στο οικείο κεφάλαιο τα οποία αναλύουν τις εκφράσεις του προσώπου και τον τόνο της φωνής όπως έχουν καταγραφεί σε βιντεοσκοπημένες συνεντεύξεις υποψήφιων εργαζομένων προκειμένου να κριθεί η καταλληλότητά των υποψηφίων για μια θέση εργασίας). Τα συστήματα περιορισμένου κινδύνου (πχ. chatbots) προβλέπεται ότι θα πρέπει να συμμορφώνονται με ελάχιστες απαιτήσεις διαφάνειας, ενώ τα συστήματα ελάχιστου κινδύνου μπορούν να αναπτύσσονται χωρίς πρόσθετες νομικές υποχρεώσεις. Τα συστήματα τεχνητής νοημοσύνης τα οποία θεωρείται ότι ελλοχεύουν «υψηλό κίνδυνο» για την υγεία, την ασφάλεια και τα θεμελιώδη δικαιώματα των φυσικών προσώπων πρέπει να πληρούν συγκεκριμένες προϋποθέσεις πριν εισέλθουν στην αγορά της ΕΕ (κατάλληλα συστήματα εκτίμησης και μετριασμού των κινδύνων,

¹³⁹ Απεικόνιση σε αναφορά της Ευρωπαϊκής Επιτροπής για τον Ευρωπαϊκό Κανονισμό για την τεχνητή νοημοσύνη, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>, τ.π. 02/04/2024.

υψηλή ποιότητα των συνόλων δεδομένων που τροφοδοτούν το σύστημα ώστε να ελαχιστοποιούνται οι κίνδυνοι και οι διακρίσεις, καταγραφή των δραστηριοτήτων με σκοπό τη διασφάλιση της ιχνηλασιμότητας των αποτελεσμάτων, λεπτομερή τεκμηρίωση που παρέχει όλες τις απαραίτητες πληροφορίες σχετικά με το σύστημα και τον σκοπό του ώστε οι αρχές να αξιολογούν τη συμμόρφωσή του, σαφείς και επαρκείς πληροφορίες για τον χρήστη, κατάλληλα μέτρα ανθρώπινης εποπτείας για την ελαχιστοποίηση του κινδύνου, υψηλό επίπεδο ανθεκτικότητας, ασφάλειας και ακρίβειας¹⁴⁰).

Στην κατηγορία «υψηλού κινδύνου» (άρθρο 6 του Κανονισμού) εντάσσονται μεταξύ άλλων και τα συστήματα που εφαρμόζονται στον τομέα της εργασίας. Τα συστήματα αυτά μπορούν, ωστόσο, υπό προϋποθέσεις που τίθενται στην παρ. 3 του αρ. 6 του Κανονισμού να μην θεωρηθούν υψηλού κινδύνου, όταν για παράδειγμα το σύστημα εντοπίζει μοτίβα λήψης αποφάσεων και δεν αποσκοπεί να αντικαταστήσει την ανθρώπινη αξιολόγηση χωρίς κατάλληλη ανθρώπινη επανεξέταση ή το σύστημα εκτελεί προπαρασκευαστικό καθήκον αξιολόγησης.

Ως συστήματα υψηλού κινδύνου στον τομέα της εργασίας αναφέρονται: α) τα συστήματα τεχνητής νοημοσύνης που προορίζονται να χρησιμοποιηθούν για την πρόσληψη ή την επιλογή φυσικών προσώπων, ιδίως για τη δημοσίευση στοχευμένων αγγελιών θέσεων εργασίας, την ανάλυση και το φιλτράρισμα των αιτήσεων εργασίας και την αξιολόγηση των υποψηφίων και β) τα συστήματα τεχνητής νοημοσύνης τα οποία προορίζονται να χρησιμοποιηθούν για τη λήψη αποφάσεων που επηρεάζουν τους όρους των εργασιακών σχέσεων, την προώθηση ή την καταγγελία συμβατικών σχέσεων που σχετίζονται με την εργασία, για την κατανομή καθηκόντων με βάση τη συμπεριφορά του ατόμου ή τα προσωπικά γνωρίσματα ή χαρακτηριστικά ή για την παρακολούθηση ή και την αξιολόγηση των επιδόσεων και της συμπεριφοράς των

¹⁴⁰«Μια Ευρώπη έτοιμη για την ψηφιακή εποχή: η Επιτροπή προτείνει νέους κανόνες και δράσεις για την αριστεία και την εμπιστοσύνη στην τεχνητή νοημοσύνη», Ιστοσελίδα Ευρωπαϊκής Επιτροπής, Δελτίο τύπου 21 Απριλίου 2021, https://ec.europa.eu/commission/presscorner/detail/el/ip_21_1682, τ.π. 02.04.2024.

προσώπων που τελούν σε τέτοιες σχέσεις. Για τα συστήματα αυτά χαρακτηριστικά αναφέρεται στην αιτιολογική σκέψη 57 του Κανονισμού ότι θα πρέπει να ταξινομηθούν ως υψηλού κινδύνου, καθώς μπορούν να επηρεάσουν σημαντικά τις μελλοντικές προοπτικές σταδιοδρομίας, τα προς το ζην, καθώς και τα δικαιώματα των εργαζομένων. Επιπλέον αναφέρεται στην ίδια ως άνω αιτιολογική σκέψη ότι οι συμβατικές σχέσεις που σχετίζονται με την εργασία θα πρέπει να καλύπτουν κατά τρόπο ουσιαστικό τους εργαζόμενους μέσω πλατφορμών και ότι τα συστήματα αυτά, μπορεί να διαιωνίσουν τις διακρίσεις λόγω φύλου, ηλικίας, αναπηρίας, φυλετικής ή εθνοτικής καταγωγής ή γενετήσιου προσανατολισμού. Στα συστήματα υψηλού κινδύνου, συνεπώς, εντάσσονται τα παραδείγματα εφαρμογών τεχνητής νοημοσύνης στις εργασιακές σχέσεις που παραθέσαμε στο οικείο κεφάλαιο της παρούσας.

Ο Κανονισμός προστατεύει τους εργαζόμενους από άδικες και μεροληπτικές αποφάσεις που λαμβάνονται από συστήματα τεχνητής νοημοσύνης μέσω διαφόρων διατάξεών του. Ειδικότερα, στο άρθρο 9 προβλέπεται ότι οι πάροχοι συστημάτων τεχνητής νοημοσύνης «υψηλού κινδύνου», δηλαδή εκείνοι που αναπτύσσουν το σύστημα, οφείλουν να εφαρμόζουν καθ' όλη τη διάρκεια του κύκλου ζωής του συστήματος τεχνητής νοημοσύνης συστήματα διαχείρισης κινδύνου για τον εντοπισμό των κινδύνων που απορρέουν από τη χρήση του συστήματος για την υγεία, την ασφάλεια ή τα θεμελιώδη δικαιώματα και να υιοθετούν μέτρα αντιμετώπισης των κινδύνων. Περαιτέρω τα συστήματα αυτά υποβάλλονται σε δοκιμή βάσει προκαθορισμένων δεικτών μέτρησης πριν τη διάθεσή τους στην αγορά για το σκοπό του προσδιορισμού των κατάλληλων και στοχευμένων μέτρων διαχείρισης κινδύνου. Έτσι εάν μια εταιρεία για παράδειγμα αναπτύσσει ένα σύστημα τεχνητής νοημοσύνης για την υποστήριξη της διαδικασίας πρόσληψη αιτούντων εργασία βάσει ανάλυσης του βιογραφικού τους, το σύστημα αυτό ταξινομείται ως σύστημα «υψηλού κινδύνου» και θα πρέπει να ελεγχθεί κατά πόσο είναι δίκαιο και εάν εισάγει ή διαιωνίζει προκαταλήψεις σε βάρος ορισμένων δημογραφικών ομάδων, βάσει φύλου, φυλής ή ηλικίας κλπ. Περαιτέρω, εφόσον εντοπιστεί κίνδυνος αθέμιτων διακρίσεων θα πρέπει αυτός ο κίνδυνος να αντιμετωπιστεί με κατάλληλα μέτρα.

Στο άρθρο 10 του Κανονισμού αναφέρεται ότι τα σύνολα δεδομένων εκπαίδευσης, επικύρωσης και δοκιμών στα συστήματα «υψηλού κινδύνου» θα πρέπει να είναι συναφή, επαρκώς αντιπροσωπευτικά και στο μέγιστο δυνατό βαθμό πλήρη και χωρίς σφάλματα έναντι του επιδιωκόμενου σκοπού. Επίσης, θα πρέπει να έχουν τις κατάλληλες στατιστικές ιδιότητες όσον αφορά τα πρόσωπα ή τις ομάδες προσώπων σε σχέση με τα οποία προορίζεται να χρησιμοποιηθεί το σύστημα υψηλού κινδύνου. Τα δεδομένα αυτά επιπρόσθετα υπόκεινται σε πρακτικές διαχείρισης δεδομένων, οι οποίες περιλαμβάνουν την εξέταση έναντι πιθανών περιπτώσεων μεροληψίας που ενδέχεται να οδηγήσουν σε διακρίσεις που απαγορεύονται βάσει του ενωσιακού δικαίου και μέτρα για την ανίχνευση, πρόληψη και μετριασμό των πιθανών περιπτώσεων μεροληψίας. Μάλιστα προκειμένου να μπορούν να ανιχνευτούν και να διορθωθούν οι περιπτώσεις μεροληψιών επιτρέπεται κατ' εξαίρεση η αναγκαία επεξεργασία ειδικών κατηγοριών προσωπικών δεδομένων υπό τον όρο κατάλληλων διασφαλίσεων για τα θεμελιώδη δικαιώματα και τις ελευθερίες των προσώπων αυτών και υπό ρητά οριζόμενες προϋποθέσεις. Η απαρίθμηση των μέτρων που θα πρέπει να τηρούνται ώστε να εξασφαλίζεται η υψηλή ποιότητα των δεδομένων έχει ως σκοπό να αποφευχθεί το φαινόμενο που συχνά αναφέρεται στην επιστήμη των υπολογιστών ως «garbage in, garbage out»¹⁴¹, δηλαδή εν προκειμένω ότι η τροφοδότηση του αλγορίθμου με κακής ποιότητας δεδομένα ή μεροληπτικά δεδομένα θα έχει και αντίστοιχο αποτέλεσμα. Οι διατυπώσεις αυτές είναι πολύ σημαντικές διότι εισάγεται σε νομοθετικό κείμενο η υποχρέωση της υψηλής ποιότητας των δεδομένων εκπαίδευσης και της αντιπροσωπευτικότητας αυτών, η οποία αποτελεί βασικό στοιχείο στη δημιουργία συστημάτων τεχνητής νοημοσύνης χωρίς προκατάληψη¹⁴², όπως έχουμε εκθέσει αναλυτικώς.

¹⁴¹ Kuśmierczyk, M. (08.05.2022), Algorithmic Bias in the Light of the GDPR and the Proposed AI Act. Maciej Olejnik, Wiktoria Morawska (ed.), "(In)equality. Faces of modern Europe", Wydawnictwo Centrum Studiów Niemieckich i Europejskich im. Willy'ego Brandta, Wrocław, 2022 (forthcoming), Available at SSRN: <https://ssrn.com/abstract=4117936> or <http://dx.doi.org/10.2139/ssrn.4117936>, τ.π. 02.04.2024.

¹⁴² Κωστή Η. Μπορεί ο αλγόριθμος να είναι δίκαιος- Ο αλγόριθμος και ο λευκός άνδρας προκατάληψης στην τεχνητή νοημοσύνη, σε συλλογικό έργο Μήτρου Λ., Μπορεί ο αλγόριθμος να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, πανεπιστημιακές εκδόσεις Κρήτης, 2023, σελ.135.

Η διαφάνεια και η ανθρώπινη εποπτεία στα συστήματα «υψηλού κινδύνου» αποτελεί το κλειδί για την αντιμετώπιση της μεροληψίας κατά τη χρήση αυτών των συστημάτων καθώς οι χρήστες θα μπορούν να λαμβάνουν ενημερωμένες αποφάσεις ενώ τα πρόσωπα που υπόκεινται στην κρίση του αλγόριθμου θα μπορούν να αμυνθούν σε περίπτωση που υποστούν δυσμενή διάκριση. Σύμφωνα, λοιπόν, με την αιτιολογική σκέψη 72 και τα άρθρα 13 και 14 του Κανονισμού τα συστήματα υψηλού κινδύνου θα πρέπει να σχεδιάζονται έτσι ώστε ο τρόπος λειτουργίας τους να είναι διαφανής και οι χρήστες των συστημάτων αυτών να μπορούν να καταλάβουν πώς λειτουργεί το σύστημα τεχνητής νοημοσύνης, να αξιολογήσουν τη λειτουργικότητά του αλλά και να κατανοήσουν τις δυνατότητες και τους περιορισμούς του. Για το σκοπό αυτό τα συστήματα αυτά θα πρέπει να συνοδεύονται από οδηγίες χρήσης σε απλή γλώσσα οι οποίες θα περιλαμβάνουν μεταξύ άλλων πληροφορίες για τις πιθανές περιστάσεις υπό τις οποίες το σύστημα υψηλού κινδύνου μπορεί να προκαλέσει κινδύνους για τα θεμελιώδη δικαιώματα, πληροφορίες για τις αλλαγές που έχουν γίνει από τον πάροχο ώστε να υπάρχει συμμόρφωση με τον Κανονισμό και για τα μέτρα ανθρώπινης εποπτείας, συμπεριλαμβανομένων των μέτρων για τη διευκόλυνση της ερμηνείας των αποτελεσμάτων του συστήματος τεχνητής νοημοσύνης από τους χρήστες. Η ανθρώπινη εποπτεία, δυνατότητα η οποία θα πρέπει να εξασφαλίζεται από τον σχεδιασμό του συστήματος, έχει ως σκοπό να προλάβει ή να ελαχιστοποιήσει τους κινδύνους για την ασφάλεια και τα θεμελιώδη δικαιώματα, όταν ένα σύστημα τεχνητής νοημοσύνης υψηλού κινδύνου επιτελεί το έργο του σύμφωνα με τον σκοπό του ή όταν γίνεται κακή χρήση αυτού. Τα φυσικά πρόσωπα που ασκούν εποπτεία θα πρέπει μεταξύ άλλων να γνωρίζουν τις ικανότητες και τους περιορισμούς του συστήματος ώστε να μπορούν να παρακολουθούν την απόδοση του και να αντιμετωπίζουν τυχόν ανωμαλίες, να ερμηνεύουν σωστά τα δεδομένα που παράγονται από το σύστημα, να έχουν τη δυνατότητα να αποφασίζουν εάν θα χρησιμοποιήσουν ή όχι τα αποτελέσματα του συστήματος και σε περίπτωση ανάγκης να μπορούν να διακόπτουν τη λειτουργία του

συστήματος. Η πρόβλεψη του ανθρώπου στον βρόγχο (Human in the loop) είναι εχέγγυο ώστε να μην πλήττονται τα θεμελιώδη δικαιώματα των ατόμων και υλοποιεί την ανθρωποκεντρική προσέγγιση σύμφωνα με την οποία η τεχνολογία θα πρέπει να βρίσκεται υπό τον έλεγχο του ανθρώπου. Έτσι για παράδειγμα, αν μια εταιρεία αναπτύσσει ένα σύστημα τεχνητής νοημοσύνης για να παρακολουθεί την παραγωγικότητα των εργαζομένων αναλύοντας δεδομένα όπως ο χρόνος που αφιερώνεται σε διάφορες εργασίες, με σκοπό να βελτιώσει τη ροή εργασίας, το σύστημα αυτό θα πρέπει να σχεδιαστεί με τρόπο που να επιτρέπει την κατανόηση και την επεξήγηση των δεδομένων εξόδου του. Επιπλέον, το σύστημα πρέπει να περιλαμβάνει μηχανισμούς ανθρώπινης επίβλεψης, που θα επιτρέπουν στους εργοδότες να επανεξετάζουν τις αξιολογήσεις της τεχνητής νοημοσύνης σε περίπτωση σφάλματος (πχ. όταν το σύστημα χαρακτηρίζει εσφαλμένα έναν υπάλληλο ως μη παραγωγικό λόγω τεχνικών προβλημάτων όπως σε περίπτωση βλάβης υπολογιστή).

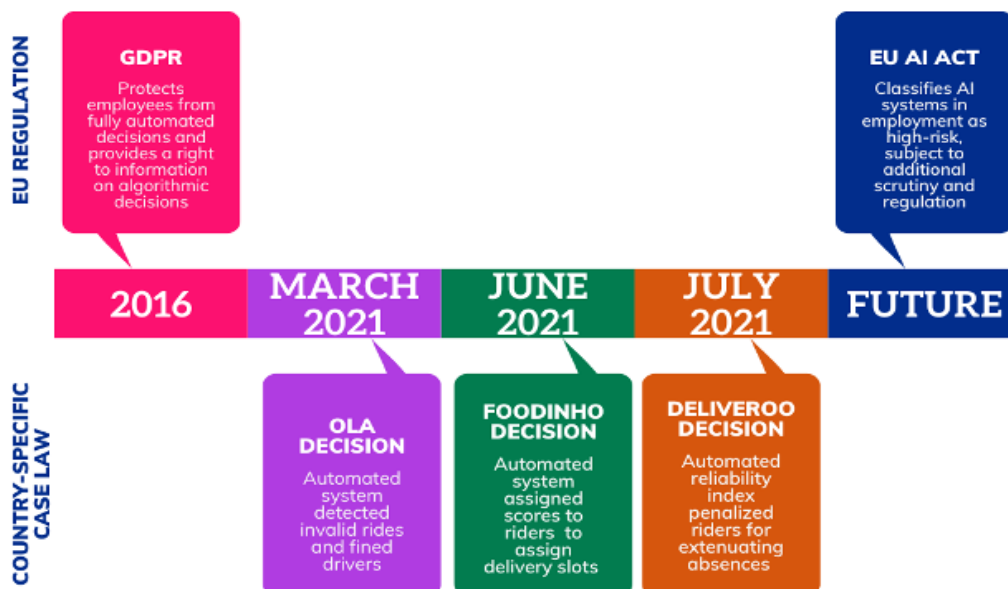
Στο άρθρο 17 προβλέπεται ότι πριν διατεθούν στην αγορά τα συστήματα τεχνητής νοημοσύνης υψηλού κινδύνου, πρέπει να υποβάλλονται σε αξιολογήσεις εφαρμόζοντας συστήματα διαχείρισης ποιότητας για να διασφαλιστεί ότι συμμορφώνονται με τις απαιτήσεις του Κανονισμού.

Το άρθρο 86 προβλέπει ότι κάθε θιγόμενο άτομο που υπόκειται σε απόφαση που λαμβάνεται με χρήση συστήματος τεχνητής νοημοσύνης υψηλού κινδύνου η οποία παράγει για αυτό έννομα αποτελέσματα ή το επηρεάζει σημαντικά με τρόπο που θεωρεί ότι έχει αρνητικό αντίκτυπο στην υγεία, την ασφάλεια ή τα θεμελιώδη δικαιώματά του θα έχει το δικαίωμα να λάβει από τον χρήστη του συστήματος σαφείς και ουσιαστικές εξηγήσεις για το ρόλο του συστήματος τεχνητής νοημοσύνης στη λήψη της απόφασης και τα κύρια στοιχεία της ληφθείσας απόφασης. Έτσι, εάν ένας υπάλληλος λάβει χαμηλή βαθμολογία απόδοσης στην εργασία από ένα σύστημα τεχνητής νοημοσύνης, ο εργαζόμενος έχει το δικαίωμα να ζητήσει εξηγήσεις για το πώς το σύστημα τεχνητής νοημοσύνης κατέληξε σε αυτήν την απόφαση. Ακολούθως

η εργοδότη εταιρεία θα πρέπει να παρέχει μια σαφή, κατανοητή επεξήγηση της διαδικασίας λήψης αποφάσεων του συστήματος, παρέχοντας λεπτομέρειες σχετικά με τα δεδομένα που χρησιμοποιούνται (π.χ. ποσοστά ολοκλήρωσης εργασιών, σχόλια από συναδέλφους) και τα συγκεκριμένα κριτήρια που το σύστημα τεχνητής νοημοσύνης θεώρησε σημαντικά ώστε κατέληξε στη συγκεκριμένη βαθμολογία απόδοσης στην εργασία. Έτσι, διασφαλίζεται η διαφάνεια και παρέχεται η δυνατότητα στον εργαζόμενο να αμφισβητήσει την απόφαση εάν πιστεύει ότι ήταν άδικη.

Στο άρθρο 26 παρ. 7 προβλέπεται ότι όταν οι χρήστες συστημάτων τεχνητής νοημοσύνης υψηλού κινδύνου είναι εργοδότες και πρόκειται να θέσουν σε λειτουργία ή να χρησιμοποιήσουν ένα τέτοιο σύστημα στο χώρο εργασίας θα πρέπει προηγουμένως να ενημερώνουν τους εκπροσώπους των εργαζομένων και τους επηρεαζόμενους εργαζόμενους ότι θα υπόκεινται στη χρήση του συστήματος τεχνητής νοημοσύνης υψηλού κινδύνου. Η διαφάνεια είναι απαραίτητη για τη διασφάλιση των συμφερόντων των εργαζομένων, λόγω της συγκέντρωσης όλης της πληροφορίας στη σφαίρα επιρροής του εργοδότη.

Από τα προαναφερθέντα προκύπτει ότι ο Κανονισμός παρέχει ένα ισχυρό πλαίσιο για την προστασία των εργαζομένων κατά τη χρήση τεχνητής νοημοσύνης στις εργασιακές σχέσεις επιβάλλοντας τη διαφάνεια, την ανθρώπινη εποπτεία και τη λογοδοσία, ώστε να διασφαλίζεται ότι τα συστήματα τεχνητής νοημοσύνης που χρησιμοποιούνται στο εργασιακό περιβάλλον είναι δίκαια, αμερόληπτα και σέβονται τα δικαιώματα των εργαζομένων. Επιπλέον, με τον Κανονισμό παρέχονται στους εργαζόμενους τα εργαλεία για να κατανοήσουν και να αμφισβητήσουν τις αλγοριθμικές αποφάσεις που τους επηρεάζουν. Η μη συμμόρφωση με τον Κανονισμό μπορεί να οδηγήσει σε σημαντικά πρόστιμα και κυρώσεις, υποχρεώνοντας τους εργοδότες να ακολουθούν θεμιτές πρακτικές.



Απεικόνιση με τη ρύθμιση της τεχνητής νοημοσύνης στην ΕΕ και νομολογία που αφορά την αλγοριθμική διοίκηση των εργασιακών σχέσεων¹⁴³

5.4. Ο Ν. 4961/2022

Τον Ιούλιο του 2022 ψηφίστηκε στην Ελλάδα ο Ν. 4961/2022 (ΦΕΚ Α 146/27.07.2022) με σκοπό να τεθούν τα θεμέλια για την ανάπτυξη θεμιτής και ασφαλούς τεχνητής νοημοσύνης, με ενισχυμένη λογοδοσία και διαφάνεια και με κατάλληλες εγγυήσεις για την προστασία των δικαιωμάτων των προσώπων που υφίστανται συνέπειες από αυτήν.

Σύμφωνα λοιπόν με το νόμο αυτό, ο οποίος δίνει έμφαση στη χρήση εφαρμογών τεχνητής νοημοσύνης από του φορείς του δημοσίου τομέα, οι φορείς του δημόσιου μπορούν κατά την άσκηση των αρμοδιοτήτων τους να χρησιμοποιούν συστήματα τεχνητής νοημοσύνης για τη λήψη ή υποστήριξη λήψης απόφασης ή την έκδοση πράξης οι οποίες επηρεάζουν τα δικαιώματα φυσικών ή νομικών προσώπων εφόσον η χρήση αυτή προβλέπεται ρητά σε ειδική διάταξη νόμου που περιλαμβάνει

¹⁴³Portillo Chavez K., Bahr J., Vartanian T., (06.12.2022), AI has made its way to the workplace. So how have laws kept pace? <https://oecd.ai/en/wonk/workplace-regulation-2022> τ.π. 02.05.2024

κατάλληλες εγγυήσεις για την προστασία των δικαιωμάτων των προσώπων και εφόσον εκπονήσουν προηγουμένως μελέτη αλγοριθμικής εκτίμησης αντικτύπου. Η μελέτη αλγοριθμικής εκτίμησης αντικτύπου περιλαμβάνει μεταξύ άλλων τον επιδιωκόμενο σκοπό, τις παραμέτρους λειτουργίας του συστήματος, το είδος των αποφάσεων που λαμβάνονται, τις κατηγορίες δεδομένων που εισάγονται στο σύστημα ή παράγονται από αυτό, τους κινδύνους που ενδέχεται να προκύψουν για τα δικαιώματα των προσώπων και το προσδοκώμενο όφελος που απορρέει για την κοινωνία σε συνάρτηση με τους ενδεχόμενους κινδύνους και επιπτώσεις που δύναται να επιφέρει η χρήση του συστήματος, ιδίως για φυλετικές, εθνοτικές, κοινωνικές ή ηλικιακές ομάδες και κατηγορίες του πληθυσμού όπως τα άτομα με αναπηρία ή χρόνιες παθήσεις. Με την επιφύλαξη των άρθρων 12 έως 14 του ΓΚΠΔ ο δημόσιος φορέας παρέχει δημόσια πληροφορίες οι οποίες περιλαμβάνουν μεταξύ άλλων τις παραμέτρους λειτουργίας του συστήματος και τις κατηγορίες των αποφάσεων που λαμβάνονται ή των πράξεων που εκδίδονται με τη συμμετοχή του συστήματος ή υποστηρίζονται από αυτό και διασφαλίζει ότι το φυσικό ή νομικό πρόσωπο το οποίο αφορά η απόφαση λαμβάνει σε απλή γλώσσα γνώση των παραμέτρων στις οποίες στηρίχθηκε η λήψη της απόφασης. Η Εθνική Αρχή Διαφάνειας (Ε.Α.Δ.), ορίζεται αρμόδια για τον χειρισμό καταγγελιών ή αναφορών, που σχετίζονται με παράβαση των υποχρεώσεων ενημέρωσης και διαφάνειας. Με την επιφύλαξη των διατάξεων για την προστασία του στρατιωτικού, εμπορικού και βιομηχανικού απορρήτου, σε κάθε προκήρυξη ανάθεσης δημόσιας σύμβασης για ανάπτυξη συστήματος τεχνητής νοημοσύνης, περιλαμβάνεται όρος για την υποχρέωση του αναδόχου να παρέχει προς τον φορέα του δημόσιου τις πληροφορίες για τις παραμέτρους λειτουργίας του συστήματος και τις κατηγορίες των αποφάσεων που λαμβάνονται και για την υποχρέωση λήψης κατάλληλων μέτρων κατά τον σχεδιασμό, ανάπτυξη και παραγωγική λειτουργία, ώστε να διασφαλίζεται η συμβατότητα του συστήματος με το νομικό πλαίσιο, ιδίως, αναφορικά με την προστασία της ανθρώπινης αξιοπρέπειας, τον σεβασμό της ιδιωτικής ζωής και την προστασία των δεδομένων προσωπικού χαρακτήρα, την απαγόρευση των διακρίσεων, την ισότητα γυναικών και ανδρών, την

ελευθερία της έκφρασης, την καθολική πρόσβαση των ατόμων με αναπηρία, τα δικαιώματα των εργαζομένων και την αρχή της χρηστής διοίκησης. Στη σύμβαση που θα συναφθεί περιλαμβάνεται ο όρος ότι το σύστημα παραδίδεται με όρους ώστε ο φορέας να μπορεί να μελετά τον τρόπο λειτουργίας του και τις παραμέτρους και να μπορεί να το βελτιώνει. Οι φορείς του δημοσίου οφείλουν να τηρούν μητρώο με τα συστήματα τεχνητής νοημοσύνης της παρ. 1 του αρ. 4 που χρησιμοποιούν, το οποίο περιλαμβάνει σχετικές πληροφορίες μεταξύ των οποίων περιλαμβάνονται και οι παράμετροι λειτουργίας και το οποίο θέτουν στη διάθεση της ΕΑΔ εφόσον ζητηθεί.

Προς τον σκοπό της παροχής εγγυήσεων για τη διασφάλιση της προστασίας των δικαιωμάτων των εργαζομένων προβλέπεται η υποχρέωση κάθε επιχείρησης του ιδιωτικού τομέα, εφόσον χρησιμοποιεί συστήματα τεχνητής νοημοσύνης, τα οποία επηρεάζουν οποιαδήποτε διαδικασία λήψης αποφάσεων σχετικά με τους εργαζομένους ή τους υποψήφιους εργαζομένους και έχει αντίκτυπο στις συνθήκες εργασίας, την επιλογή, την πρόσληψη ή την αξιολόγησή τους, σε κάθε περίπτωση πριν την πρώτη χρήση τους, να παρέχει επαρκή και σαφή πληροφόρηση σε κάθε εργαζόμενο ή υποψήφιο εργαζόμενο, η οποία περιλαμβάνει κατ' ελάχιστον τις παραμέτρους στις οποίες στηρίζεται η λήψη της απόφασης, με την επιφύλαξη των περιπτώσεων που προϋποθέτουν προηγούμενη ενημέρωση και διαβούλευση και διασφαλίζει την τήρηση της αρχής της ίσης μεταχείρισης και της καταπολέμησης των διακρίσεων στην απασχόληση και την εργασία λόγω φύλου, φυλής, χρώματος, εθνικής ή εθνοτικής καταγωγής γενεαλογικών καταβολών, θρησκευτικών ή άλλων πεποιθήσεων, αναπηρίας ή χρόνιας πάθησης, ηλικίας, οικογενειακής ή κοινωνικής κατάστασης, σεξουαλικού προσανατολισμού, ταυτότητας ή χαρακτηριστικών φύλου. Η υποχρέωση αυτή καταλαμβάνει και τις ψηφιακές πλατφόρμες κατά την έννοια του άρθρου 68 του ν. 4808/2021, ως προς τα φυσικά πρόσωπα που συνδέονται μαζί τους με συμβάσεις εξαρτημένης εργασίας ή ανεξάρτητων υπηρεσιών ή έργου. Περαιτέρω τονίζεται ότι δεν θίγεται το άρθρο 22 ΓΚΠΔ για την αυτοματοποιημένη ατομική λήψη αποφάσεων και την κατάρτιση προφίλ. Σε περίπτωση παραβίασης της υποχρέωσης επιβάλλονται οι διοικητικές και ποινικές κυρώσεις των άρθρων 24 και 28, αντίστοιχα,

του ν. 3996/2011. Αρμόδια αρχή για την επιβολή των διοικητικών κυρώσεων του πρώτου εδαφίου είναι το Σώμα Επιθεώρησης Εργασίας (Σ.Ε.Π.Ε.). Τέλος, προκειμένου να εξασφαλιστεί η λογοδοσία προβλέπεται η υποχρέωση των επιχειρήσεων του ιδιωτικού τομέα, οι οποίες συνιστούν μεσαία ή μεγάλη οντότητα (παρ. 5 και 6 αρ. 2 Ν. 4308/2014) να τηρούν ηλεκτρονικό μητρώο εφαρμογών τεχνητής νοημοσύνης που χρησιμοποιούνται στο πλαίσιο κατάρτισης προφίλ καταναλωτών ή για την αξιολόγηση εργαζομένων, με σχετικές πληροφορίες και προβλέπεται η υποχρέωση τους να καταρτίζουν και να τηρούν πολιτική δεοντολογικής χρήσης δεδομένων, η οποία περιλαμβάνει πληροφορίες σχετικά με τα μέτρα, τις ενέργειες και τις διαδικασίες που εφαρμόζει σε θέματα δεοντολογίας δεδομένων κατά τη χρήση συστημάτων τεχνητής νοημοσύνης. Στο μητρώο περιλαμβάνονται για κάθε σύστημα, κατ' ελάχιστον, οι εξής πληροφορίες: α) η περιγραφή των παραμέτρων λειτουργίας, των δυνατοτήτων και των τεχνικών χαρακτηριστικών του συστήματος, β) ο αριθμός και η ιδιότητα των φυσικών προσώπων που αφορά ή ενδέχεται να αφορά, γ) οι τεχνικές πληροφορίες που αφορούν στον προμηθευτή ή εξωτερικούς συνεργάτες που συμμετέχουν στην ανάπτυξη ή λειτουργία του συστήματος, δ) η περίοδος λειτουργίας του συστήματος, και ε) τα μέτρα που λαμβάνονται για την ασφαλή λειτουργία τους.

Ο Ν. 4961/2022 αποτελεί μια ιδιαίτερα σημαντική πρωτοβουλία του Έλληνα νομοθέτη, η οποία ήρθε πριν την ψήφιση και θέση σε εφαρμογή του Κανονισμού για την τεχνητή νοημοσύνη και σε χρονικό διάστημα κατά το οποίο το κείμενο της οδηγίας για τη βελτίωση των συνθηκών εργασίας των εργαζομένων σε πλατφόρμες, ήταν ακόμα υπό διαπραγμάτευση. Ο νόμος διαγράφει το πλαίσιο ανάπτυξης των συστημάτων τεχνητής νοημοσύνης με γνώμονα την ηθική, τη διαφάνεια και τη λογοδοσία, ενώ παρέχει σημαντική προστασία στους εργαζόμενους που υποβάλλονται σε πρακτικές αλγοριθμικής διοίκησης των εργασιακών τους σχέσεων, καθιερώνοντας την υποχρέωση ενημέρωσής τους για τις πρακτικές αυτές και υποχρεώνοντας τους εργοδότες να τηρούν την αρχή της ισότητας μη διάκρισης κατά τη χρήση αλγορίθμων, επιβάλλοντας κυρώσεις σε αυτούς σε περίπτωση μη συμμόρφωσης με τις υποχρεώσεις αυτές. Έτσι, καθίσταται σαφές ότι οι εργοδότες δεν

μπορούν να αποποιηθούν την ευθύνη σε περίπτωση που χρησιμοποιούν συστήματα τεχνητής νοημοσύνης για τη διοίκηση των εργασιακών σχέσεων. Επιπλέον, γίνεται μνεία στο άρθρο 22 του ΓΚΠΔ ώστε τονίζεται η απαγόρευση αξιοποίησης δεδομένων ειδικών κατηγοριών του αρ. 9 ΓΚΠΔ για τη λήψη αλγοριθμικών αποφάσεων, το δικαίωμα στην εξασφάλιση ανθρώπινης παρέμβασης από την πλευρά του υπεύθυνου επεξεργασίας, έκφρασης άποψης και αμφισβήτηση της απόφασης, όταν λαμβάνεται στο πλαίσιο σύναψης ή εκτέλεσης της εργασιακής σύμβασης μια πλήρως αυτοματοποιημένη απόφαση συμπεριλαμβανομένης της κατάρτισης προφίλ η οποία παράγει έννομα αποτελέσματα ή επηρεάζει σημαντικά με παρόμοιο τρόπο τον εργαζόμενο. Έτσι, επιτυγχάνεται η προστασία των εργαζομένων και αποτρέπονται οι λανθασμένες ή άδικες αποφάσεις. Φυσικά ισχύουν εκ του νόμου και τα υπόλοιπα δικαιώματα του ΓΚΠΔ, όπως για παράδειγμα το δικαίωμα πρόσβασης (άρθρο 15 ΓΚΠΔ) με ιδιαίτερα χρήσιμη την πρόβλεψη στην περίπτωση η' της παρ. 1 αυτού για παροχή πληροφοριών για τη λογική του αλγόριθμου στις αυτοματοποιημένες αποφάσεις και τις επιπτώσεις αυτής της επεξεργασίας και ακολούθως το δικαίωμα διόρθωσης (άρθρο 16 ΓΚΠΔ) των εσφαλμένων δεδομένων. Έτσι, ένας εργαζόμενος μπορεί να ζητήσει πληροφορίες σχετικά με το πώς ένας αλγόριθμος αξιολογεί την απόδοσή του, να μάθει τη λογική του αλγόριθμου ποια δεδομένα αξιοποιεί για την απόφαση, να ζητήσει τη διόρθωση αυτών αν είναι εσφαλμένα, να ζητήσει ανθρώπινη παρέμβαση και να ακουστεί και να αμφισβητήσει την απόφαση.

Η υποχρέωση για διαφάνεια διαπνέει τον Ν.4961/2022. Η απαίτηση για διαφάνεια και ενημέρωση των εργαζομένων που υπόκεινται στην κρίση αλγορίθμων (αρ. 9 παρ. 1 Ν. 4961/2022) είναι σημαντική για την αποφυγή καταχρηστικών πρακτικών, ωστόσο ανακύπτει ο προβληματισμός κατά πόσο οι εργαζόμενοι έχουν τα κατάλληλα εργαλεία και γνώση για να κατανοήσουν πλήρως τις πληροφορίες που θα τους παρέχονται. Ενδεχομένως σε αυτό το σημείο σημαντικός να είναι ο ρόλος των ενώσεων των εργαζομένων οι οποίες θα μπορούσαν να συνδράμουν προς αυτή την κατεύθυνση με τη βοήθεια ειδικών. Περαιτέρω, η ίδια διάταξη δεν εξειδικεύει τι θα πρέπει να περιλαμβάνει η ενημέρωση αυτή ούτε τον τρόπο με τον οποίο πρέπει να

γίνεται και περιορίζεται στην αναφορά ότι η ενημέρωση θα πρέπει να περιλαμβάνει κατ' ελάχιστον τις παραμέτρους που ελήφθησαν υπόψη για τη λήψη της απόφασης. Συνεπώς, θα πρέπει να συμπληρώνεται η διάταξη με την συνεφαρμογή του ΓΚΠΔ. Επιπλέον στην ίδια διάταξη προβλέπεται ότι οι εργοδότες θα πρέπει να διασφαλίζουν την τήρηση της αρχής της ίσης μεταχείρισης και της καταπολέμησης των διακρίσεων στην εργασία δεν αναφέρεται, όμως, ποια μέτρα θα πρέπει να λάβουν οι εργοδότες προκειμένου να το επιτύχουν αυτό ή πως θα αξιολογούν και θα προσαρμόζουν τους αλγόριθμους για την αποφυγή τέτοιων φαινομένων. Η ανάθεση της αρμοδιότητας επιβολής των διοικητικών κυρώσεων σε περίπτωση παραβίασης από τους εργοδότες της υποχρέωσης του αρ. 9 παρ. 1 στο Σ.Ε.Π.Ε. είναι προς την ορθή κατεύθυνση μιας και στις αρμοδιότητές του ανήκει ο έλεγχος της εφαρμογής των διατάξεων της εργατικής νομοθεσίας και συνεπώς στο πλαίσιο αυτό εποπτεύει την εφαρμογή της νομοθεσίας για την προώθηση της αρχής της ίσης μεταχείρισης, έχοντας τη δυνατότητα να επιβάλλει στους παραβάτες διοικητικές κυρώσεις ή να προσφεύγει στη δικαιοσύνη για επιβολή ποινικών κυρώσεων. Ωστόσο, προκειμένου να μπορέσει να ανταποκριθεί στον ρόλο αυτό το ΣΕΠΕ θα πρέπει να στελεχωθεί με εξειδικευμένο προσωπικό όπως αναλυτές δεδομένων, νομικούς και κοινωνιολόγους, ώστε να μπορεί επί της ουσίας να ελέγξει τα συστήματα τεχνητής νοημοσύνης που χρησιμοποιούνται και να κρίνει εάν αυτά είναι δίκαια.

Το δημιουργούμενο κανονιστικό πλαίσιο προσπαθεί να συμβάλει στην αποτροπή των όποιων δυσμενών συνεπειών ενδέχεται να προκύψουν από τη χρήση των αναδυόμενων τεχνολογιών, όπως είναι η τεχνητή νοημοσύνη, διατηρώντας παράλληλα έναν ευέλικτο χαρακτήρα, προκειμένου οι θεσπιζόμενες ρυθμίσεις να προσαρμόζονται διαρκώς στις τεχνολογικές εξελίξεις και να μην καθίστανται ανεπίκαιρες σε μικρό χρονικό διάστημα. Με επίκεντρο των ρυθμίσεων την προστασία των θεμελιωδών δικαιωμάτων των προσώπων που εμπλέκονται με οποιονδήποτε τρόπο στη χρήση τους, ο Ν. 4961/2022 θέτει όρια στην ανεξέλεγκτη χρήση των τεχνολογιών που επιλέγει να ρυθμίσει, προκειμένου η χρήση τους να αποβαίνει επ' ωφελεία των χρηστών τους. Το αν πράγματι οι επιλεγείσες ρυθμίσεις θα αποδειχθούν

αποτελεσματικές στην πράξη και το αν θα εκπληρωθεί ο σκοπός της θέσπισής τους, μένει να αποδειχθεί¹⁴⁴.

5.5. Η Οδηγία για τη βελτίωση των συνθηκών εργασίας των εργαζομένων σε πλατφόρμες

Όπως αναπτύξαμε στο οικείο κεφάλαιο η εργασία μέσω πλατφόρμας δημιουργεί προκλήσεις όσον αφορά την προστασία της ιδιωτικότητας και των εργασιακών δικαιωμάτων. Οι εργαζόμενοι συχνά υφίστανται δυσμενείς διακρίσεις που οφείλονται στους αλγόριθμους που οργανώνουν την εργασία στις πλατφόρμες, ενώ απουσιάζει η διαφάνεια, η ανθρώπινη εποπτεία και η λογοδοσία. Για το λόγο αυτό τα θεσμικά όργανα της ΕΕ προκειμένου να βελτιώσουν τις συνθήκες εργασίας των εργαζομένων στις πλατφόρμες εξέδωσαν πρόταση οδηγίας στις 09.12.2021¹⁴⁵. Στις 14 Οκτωβρίου 2024 η Οδηγία «για τη βελτίωση των συνθηκών εργασίας των εργαζομένων σε πλατφόρμες»¹⁴⁶ υιοθετήθηκε επίσημα σε πρώτη ανάγνωση από το Συμβούλιο της ΕΕ και επί του παρόντος εκκρεμεί προς υπογραφή από το Συμβούλιο και το Ευρωπαϊκό Κοινοβούλιο. Μετά την υπογραφή της θα δημοσιευθεί στην Επίσημη Εφημερίδα της Ευρωπαϊκής Ένωσης και θα τεθεί σε ισχύ 20 ημέρες μετά. Ακολούθως, τα κράτη μέλη θα έχουν στη διάθεσή τους δύο έτη για να ενσωματώσουν τις διατάξεις της Οδηγίας στο εθνικό τους δίκαιο.

Σύμφωνα με την Οδηγία, λοιπόν, ως ψηφιακή πλατφόρμα νοείται το φυσικό ή νομικό πρόσωπο που παρέχει υπηρεσία εξ αποστάσεως με ηλεκτρονικά μέσα (πχ. ιστότοπος, εφαρμογή για κινητά τηλέφωνα), κατόπιν αιτήματος του αποδέκτη υπηρεσίας, οργανώνοντας την εργασία που εκτελείται από πρόσωπα έναντι πληρωμής και

¹⁴⁴ Κουσουνή – Πανταζοπούλου Α. (30.08.2023), Οι αναδυόμενες τεχνολογίες υπό το πρίσμα του Ν. 4961/2022, Άρθρο στο Μηνιαίο Ενημερωτικό Δελτίο ΣΕΒ – Ρυθμιστικό Περιβάλλον και Επιχειρήσεις, https://www.sev.org.gr/arthografia_mme/oi-anadyomenes-technologies-ypo-to-prisma-tou-n-4961-2022/

¹⁴⁵ Πρόταση οδηγίας του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου σχετικά με τη βελτίωση των συνθηκών εργασίας στην εργασία σε πλατφόρμες, COM (2021) 762 final της 9^{ης} Δεκεμβρίου 2021.

¹⁴⁶ European Council, Directive on Improving Working Conditions in Platform Work (PE-89-2024-INIT), Council of the European Union, 2024, <https://data.consilium.europa.eu/doc/document/PE-89-2024-INIT/en/pdf>

χρησιμοποιώντας αυτοματοποιημένα συστήματα παρακολούθησης ή λήψης αποφάσεων. Στόχος της οδηγίας είναι να διασφαλίσει ότι οι εργαζόμενοι στις πλατφόρμες θα υπάγονται στο σωστό νομικό καθεστώς που ανταποκρίνεται στην πραγματική σχέση εργασίας τους με την πλατφόρμα, κάθε φορά που διαπιστώνονται πραγματικά περιστατικά που υποδεικνύουν διεύθυνση και έλεγχο. Τα κράτη μέλη για το σκοπό αυτό θα πρέπει να θεσπίζουν αποτελεσματικό ανατρέψιμο νομικό τεκμήριο σχέσης εργασίας, το οποίο θα εφαρμόζεται σε όλες τις σχετικές διοικητικές ή δικαστικές διαδικασίες. Περαιτέρω, με την οδηγία επιδιώκεται η αύξηση της διαφάνειας, της δικαιοσύνης, της ανθρώπινης εποπτείας, της ασφάλειας και της λογοδοσίας στην αλγοριθμική διοίκηση στην εργασία σε πλατφόρμα.

Οι πλατφόρμες απαγορεύεται μέσω αυτοματοποιημένων συστημάτων παρακολούθησης ή λήψης αποφάσεων να επεξεργάζονται προσωπικά δεδομένα που σχετίζονται με τη συναισθηματική ή ψυχολογική κατάσταση των εργαζομένων ή που αφορούν σε ιδιωτικές συζητήσεις αυτών ή προσωπικά δεδομένα που συλλέγονται εκτός ωραρίου εργασίας. Επιπλέον, δεν μπορούν να επεξεργάζονται προσωπικά δεδομένα για την εξαγωγή συμπερασμάτων σχετικά με τη φυλετική ή εθνοτική καταγωγή, το μεταναστευτικό καθεστώς, τα πολιτικά φρονήματα, τις θρησκευτικές ή φιλοσοφικές πεποιθήσεις, την αναπηρία, την κατάσταση της υγείας, συμπεριλαμβανομένης της χρόνιας νόσου ή κατάσταση HIV, τη συναισθηματική ή ψυχολογική κατάσταση, τη συμμετοχή σε συνδικαλιστική οργάνωση, τη σεξουαλική ζωή ή τον γενετήσιο προσανατολισμό ενός προσώπου. Δεν μπορούν, επιπλέον, οι πλατφόρμες να επεξεργάζονται βιομετρικά δεδομένα για την εξακρίβωση της ταυτότητας των εργαζομένων αντιπαραβάλλοντας αυτά με αποθηκευμένα βιομετρικά δεδομένα φυσικών προσώπων σε βάση δεδομένων. Οι πλατφόρμες ως υπεύθυνοι επεξεργασίας πρέπει να εκπονούν ΕΑΠΔ κατ' αρ. 35 ΓΚΠΔ, ζητώντας και την άποψη των εργαζομένων και των εκπροσώπων τους. Οι ψηφιακές πλατφόρμες εργασίας θα πρέπει να ενημερώνουν τους εργαζομένους γραπτώς και σε απλή γλώσσα το αργότερο την πρώτη εργάσιμη ημέρα σχετικά με τα αυτοματοποιημένα συστήματα που χρησιμοποιούνται για την παρακολούθηση, την επίβλεψη ή την

αξιολόγηση των εργασιακών επιδόσεων και τα αυτοματοποιημένα συστήματα λήψης αποφάσεων. Η ενημέρωση σε σχέση με τα αυτοματοποιημένα συστήματα λήψης αποφάσεων περιλαμβάνει το γεγονός ότι πρόκειται να χρησιμοποιηθούν τέτοια συστήματα, τις κατηγορίες των αποφάσεων που λαμβάνονται ή υποστηρίζονται από τα συστήματα αυτά, τις κατηγορίες των δεδομένων και τις κύριες παραμέτρους που λαμβάνουν υπόψη και τη σχετική σημασία των παραμέτρων στο πλαίσιο της αυτοματοποιημένης λήψης αποφάσεων, τους λόγους για τους οποίους λαμβάνονται αποφάσεις περιορισμού, αναστολής ή παύσης του λογαριασμού του εργαζομένου, αποφάσεις άρνησης καταβολής αμοιβής και αποφάσεις για το συμβατικό καθεστώς του εργαζομένου. Οι πληροφορίες αυτές παρέχονται και στους εκπροσώπους των εργαζομένων. Οι εργαζόμενοι έχουν το δικαίωμα φορητότητας των δεδομένων τους, δηλαδή λήψης αντιγράφου σε αναγνώσιμο από μηχάνημα υλικό φορέα, στα δεδομένα αυτά συμπεριλαμβάνονται οι βαθμολογήσεις και αξιολογήσεις των αποδεκτών των υπηρεσιών. Οι πλατφόρμες πρέπει να εποπτεύουν και να αξιολογούν τον αντίκτυπο των αποφάσεων που λαμβάνονται ή στηρίζονται από αυτοματοποιημένα συστήματα λήψης αποφάσεων (συνθήκες εργασίας, ίση μεταχείριση στην εργασία) αναθέτοντας το σχετικό καθήκον σε πρόσωπα που διαθέτουν σχετική κατάρτιση, εξουσία και ανεξαρτησία και τα οποία προστατεύονται από απόλυση ή πειθαρχικά μέτρα. Αν κατά την εποπτεία ή αξιολόγηση βρεθεί υψηλός κίνδυνος διακρίσεων στην εργασία κατά τη χρήση των αυτοματοποιημένων συστημάτων, τότε τροποποιείται ή παύει να χρησιμοποιείται το σύστημα. Προκειμένου να διασφαλιστεί η αμεροληψία και η λογοδοσία για τις αποφάσεις που λαμβάνονται ή υποστηρίζονται από αυτοματοποιημένα συστήματα, προβλέπεται η δημιουργία κατάλληλων διαύλων ώστε να μπορεί να ζητηθεί η επανεξέταση αυτών των αποφάσεων με πρόσωπα υπεύθυνους επικοινωνίας. Οι πλατφόρμες θα πρέπει να παρέχουν χωρίς αδικαιολόγητη καθυστέρηση γραπτή αιτιολογία για οποιαδήποτε απόφαση που λαμβάνεται ή υποστηρίζεται από αυτοματοποιημένο σύστημα λήψης αποφάσεων για τον περιορισμό, αναστολής ή τερματισμού του λογαριασμού του εργαζομένου πλατφόρμας, άρνησης αμοιβής για εργασία που εκτελεί ο εργαζόμενος

πλατφόρμας ή επηρεασμού της συμβατικής κατάστασης του εργαζόμενου στην πλατφόρμα. Για τις υπόλοιπες αποφάσεις παρέχονται γραπτές ή προφορικές εξηγήσεις. Οι εργαζόμενοι έχουν δικαίωμα να ζητήσουν την επανεξέταση των αποφάσεων και οι πλατφόρμες έχουν την υποχρέωση να απαντήσουν τεκμηριωμένα το αργότερο εντός δύο βδομάδων από την παραλαβή του αιτήματος. Οι πλατφόρμες πρέπει να διορθώσουν την απόφαση χωρίς καθυστέρηση αν αυτή παραβιάζει τα δικαιώματα του εργαζόμενου ή, εάν αυτό δεν είναι πλέον δυνατό, να παράσχουν επαρκή αποζημίωση. Η οδηγία προωθεί τον κοινωνικό διάλογο σχετικά με τα συστήματα αλγοριθμικής διοίκησης μέσω της καθιέρωσης συλλογικών δικαιωμάτων ενημέρωσης και διαβούλευσης για την εισαγωγή και τις ουσιαστικές αλλαγές που σχετίζονται με τη χρήση αυτοματοποιημένων συστημάτων λήψης αποφάσεων. Οι εκπρόσωποι των εργαζομένων μπορούν να επικουρούνται από εμπειρογνώμονα εφόσον είναι αναγκαίο για να εξετάσουν θέμα που είναι αντικείμενο της διαβούλευσης και να διατυπώνουν γνώμη. Όταν η πλατφόρμα έχει πάνω από 250 εργαζόμενους την δαπάνη του εμπειρογνώμονα αναλαμβάνει η πλατφόρμα. Τα κράτη μέλη πρέπει να διασφαλίζουν ότι οι εργαζόμενοι σε πλατφόρμες έχουν πρόσβαση σε έγκαιρη, αποτελεσματική και αμερόληπτη επίλυση διαφορών και δικαίωμα επανόρθωσης, συμπεριλαμβανομένης της καταβολής επαρκούς αποζημίωσης για τη ζημία που υπέστησαν σε περίπτωση παραβίασης των δικαιωμάτων τους που απορρέουν από την οδηγία. Επιπλέον, διασφαλίζουν ότι τα εθνικά δικαστήρια ή οι αρμόδιες εθνικές αρχές είναι σε θέση να διατάσσουν την πλατφόρμα να γνωστοποιήσει κάθε σχετικό στοιχείο που βρίσκεται υπό τον έλεγχό της, ακόμα και εμπιστευτικές πληροφορίες εφόσον το κρίνουν σκόπιμο στο πλαίσιο της διαδικασίας. Οι εργαζόμενοι δεν μπορούν να απολυθούν για το λόγο ότι άσκησαν τα δικαιώματά τους που απορρέουν από την οδηγία.

Ως εκ τούτου, οι εργαζόμενοι μέσω πλατφορμών και οι εκπρόσωποί τους θα απολαμβάνουν καλύτερη διαφάνεια και θα έχουν καλύτερη κατανόηση των πρακτικών αλγοριθμικής διαχείρισης, καθώς και αποτελεσματικότερη πρόσβαση σε μέσα έννομης προστασίας κατά αυτοματοποιημένων αποφάσεων. Τα δικαιώματα

αυτά θα επεκτείνουν τις υφιστάμενες διασφαλίσεις που παρέχει ο Ν. 4961/2022, ο οποίος αναφέρεται ρητά στους εργαζόμενους μέσω πλατφόρμας, ο ΓΚΠΔ όσον αφορά την επεξεργασία δεδομένων προσωπικού χαρακτήρα από αυτοματοποιημένα συστήματα λήψης αποφάσεων, καθώς και τις υποχρεώσεις για τους παρόχους και τους χρήστες συστημάτων τεχνητής νοημοσύνης όσον αφορά τη διαφάνεια και την ανθρώπινη εποπτεία των συστημάτων τεχνητής νοημοσύνης υψηλού κινδύνου στον Κανονισμό για την τεχνητή νοημοσύνη¹⁴⁷.

¹⁴⁷Πρόταση Οδηγία του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου για τη βελτίωση των συνθηκών εργασίας των εργαζομένων σε πλατφόρμες COM/2021/762 final, <https://eur-lex.europa.eu/legal-content/EL/ALL/?uri=CELEX:52021PC0762>, τ.π. 02.04.2024.

ΣΥΜΠΕΡΑΣΜΑ

Στην εποχή της παγκοσμιοποίησης, οι επιχειρήσεις αντιμετωπίζουν αυξημένο διεθνή ανταγωνισμό, προσπαθώντας να αποκτήσουν ανταγωνιστικό πλεονέκτημα για να εξασφαλίσουν την επιβίωσή τους στην αγορά. Σε αυτή την προσπάθεια, η τεχνολογία, και ιδιαίτερα η τεχνητή νοημοσύνη, αποτελεί έναν πολύτιμο σύμμαχο. Η τεχνητή νοημοσύνη μετασχηματίζει τον κόσμο της εργασίας, προσφέροντας σημαντικά οφέλη, όμως, παράλληλα ενέχει και σοβαρούς κινδύνους, όπως αυτός της αλγοριθμικής μεροληψίας. Το ζήτημα της εξασφάλισης της δικαιοσύνης στους αλγορίθμους της τεχνητής νοημοσύνης είναι εξαιρετικά κρίσιμο όταν οι αποφάσεις και συστάσεις αυτών εφαρμόζονται σε ευαίσθητους τομείς όπως αυτός της εργασίας.

Η διαφάνεια στον τρόπο λειτουργίας των αλγορίθμων και η επεξηγησιμότητα των αποτελεσμάτων τους, αποτελεί εχέγγυο τήρησης της δικαιοσύνης, καθώς επιτρέπει στους εργαζόμενους να μπορούν να αμφισβητούν τις αλγοριθμικές αποφάσεις που καθορίζουν τις συνθήκες εργασίας και το εισόδημά τους. Επιπλέον, η διαφάνεια είναι αλληλένδετη με το καθήκον λογοδοσίας της εργοδοσίας. Μια μηχανή δεν θα πρέπει μόνη χωρίς ανθρώπινη παρέμβαση και ουσιαστική εποπτεία να λαμβάνει αποφάσεις με συνέπειες κρίσιμες για τους ανθρώπους. Οι προβλέψεις και συστάσεις που εισφέρουν οι αλγόριθμοι, θα πρέπει να αποτελούν βοήθεια στη λήψη αποφάσεων και στην ανθρώπινη δράση, χωρίς να οδηγούμαστε σε απευθυνοποίηση του ανθρώπου και σε απώλεια της αυτονομίας του¹⁴⁸. Ο άνθρωπος, ως ο τελικός υπεύθυνος λήψης των αποφάσεων, πρέπει να είναι σε θέση να κατανοεί, να ελέγχει και να αιτιολογεί τις αλγοριθμικές συστάσεις που υιοθετεί, αναλαμβάνοντας πλήρως την ευθύνη για τη νομιμότητα και τη δικαιοσύνη τους. Το ισχύον νομοθετικό πλαίσιο παρέχει σημαντικές εγγυήσεις για τη διαφάνεια, τη δικαιοσύνη, την ανθρώπινη εποπτεία και

¹⁴⁸ Παπαδημητρίου Κ., Νέες τεχνολογίες, αλγόριθμοι και εργατικό δίκαιο, Εισήγηση στην ΕΣΔΙ, https://www.esdi.gr/wp-content/uploads/2022/seminars/03/dikaio-plir/papadimitriou_2022.pdf, τ.π. 31.08.2024.

τη λογοδοσία. Οι εργοδότες, υπό την απειλή κυρώσεων, υποχρεούνται πλέον να εφαρμόζουν όχι μόνο αποτελεσματικά, αλλά και δίκαια και διαφανή συστήματα.

Ωστόσο, οι προκλήσεις που απορρέουν από τη χρήση αλγορίθμων μηχανικής μάθησης παραμένουν. Η τεχνολογία εξελίσσεται ραγδαία, ενώ η νομοθεσία ακολουθεί με βραδύτερους ρυθμούς, ώστε δοκιμάζεται συχνά η αποτελεσματικότητά της. Για να ανταποκριθεί η νομοθεσία στις τεχνολογικές εξελίξεις, απαιτείται στενή συνεργασία μεταξύ νομοθετών και ειδικών στην τεχνητή νοημοσύνη. Η επιτήρηση και η αξιολόγηση των επιπτώσεων της τεχνολογίας πρέπει να είναι συνεχείς, καθώς οι αλγόριθμοι εξελίσσονται και ενδέχεται να αναπτύξουν προκαταλήψεις μετά την αρχική τους δημιουργία. Η διαφάνεια, θεμελιώδης για τη διαπίστωση της δικαιοσύνης, δεν είναι πάντοτε τεχνικά εφικτή λόγω της πολυπλοκότητας ορισμένων αλγορίθμων. Η σύγκλιση της έννοιας της δικαιοσύνης κατά το δίκαιο και την επιστήμη των υπολογιστών δεν έχει ακόμα επιτευχθεί. Η καθιέρωση κοινά αποδεκτών και ομοιόμορφα εφαρμοζόμενων μεθοδολογιών που θα διασφαλίζουν ότι τα αλγοριθμικά συστήματα λήψης αποφάσεων λειτουργούν με τρόπο που αποτρέπει τις διακρίσεις, ενισχύει τη διαφάνεια, τηρεί τις νομικές και ηθικές αρχές και προάγει τη λογοδοσία είναι απαραίτητη. Περαιτέρω, η αποτελεσματική επιβολή της νομοθεσίας απαιτεί την ενίσχυση των εποπτικών αρχών, όπως το ΣΕΠΕ και η ΑΠΔΠΧ, με επαρκείς πόρους και εξειδικευμένο επιστημονικό προσωπικό. Οι δικαστές, ως εγγυητές της δικαιοσύνης, θα κληθούν να προσαρμόσουν τις παραδοσιακές αρχές του εργατικού δικαίου στις προκλήσεις της αλγοριθμικής εποχής.

Σε ένα περιβάλλον υπεραισιοδοξίας και ενθουσιασμού για αυτήν την πολλά υποσχόμενη τεχνολογία, θα πρέπει να έχουμε κατά νου ότι η τεχνητή νοημοσύνη δεν αποτελεί πανάκεια, αποτελεί ένα ισχυρό εργαλείο που πρέπει να αξιολογείται ως προς τη χρησιμότητα και την αναγκαιότητά του σε συνάρτηση με τους κινδύνους που ενέχει. Σε κάθε περίπτωση ο άνθρωπος θα πρέπει να κρατά τα ηνία της τεχνητής

νοημοσύνης έχοντας ως προορισμό την ασφαλή μετάβαση προς την ανθρώπινη ευημερία.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ι. ΕΛΛΗΝΙΚΗ

Aloisi A., Από μηχανής αφεντικό: Οι εργοδοτικές εξουσίες στους χώρους εργασίας που διευθύνονται από αλγορίθμους και τεχνητή νοημοσύνη, ΕΕργΔ, τόμος 82ος, τεύχος 6, 2023.

Ανεξάρτητη Ομάδα Εμπειρογνομώνων Υψηλού Επιπέδου για την Τεχνητή Νοημοσύνη, συσταθείσα από την Ευρωπαϊκή Επιτροπή τον Ιούνιο του 2018, Κατευθυντήριες γραμμές δεοντολογίας για αξιόπιστη τεχνητή νοημοσύνη, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>, τ.π. 27.03.2024.

Κουσουνη – Πανταζοπούλου Α., Νομικές διαστάσεις της τεχνητής νοημοσύνης - παρόν και μέλλον, ΕλλΔνη 1/2019, σελ. 313.

Ομάδα εργασίας για την προστασία δεδομένων του άρθρου 29 της Οδηγίας 95/46/ΕΚ, Κατευθυντήριες γραμμές για την αυτοματοποιημένη λήψη αποφάσεων και την κατάρτιση προφίλ για τους σκοπούς του Κανονισμού 2016/679 με ημερομηνία έγκρισης από το Ευρωπαϊκό Συμβούλιο Προστασίας Δεδομένων την 3η Οκτωβρίου 2017 και αναθεώρησης και έκδοσης την 6η Φεβρουαρίου 2018, https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/automated-decision-making-and-profiling_el, τ.π. 31.08.2024.

Μήτρου Λ., Μπορεί ο αλγόριθμος... να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει και να διοικεί, Πανεπιστημιακές εκδόσεις Κρήτης, Ηράκλειο Κρήτης, 2023.

Παπαγεωργίου Ι., Συστήματα τεχνητής νοημοσύνης και διακριτική μεταχείριση κατά το προσυμβατικό στάδιο σύναψης της σύμβασης εργασίας, ΕφΑΔΠολΔ, 10/2022, σελ. 1125-1140.

Παπαδημητρίου Κ., Γενική Εισαγωγή σε Πληροφορική και Εργατικό Δίκαιο, Νομική Βιβλιοθήκη, Αθήνα, 2018, σελ. 5 επ.

Παπαδημητρίου Κ., (2022) Νέες τεχνολογίες, αλγόριθμοι και εργατικό δίκαιο, Εισήγηση στην ΕΣΔΙ, https://www.esdi.gr/wp-content/uploads/2022/seminars/03/dikaio-plir/papadimitriou_2022.pdf, τ.π. 31.08.2024.

Συνήγορος του Πολίτη (2018), Οδηγός Ίσης Μεταχείρισης – Εγχειρίδιο για δημοσίους υπαλλήλους από τον Συνήγορο του Πολίτη, <https://old.synigoros.gr/?i=equality.el.ifleaflet>, τ.π. 23.04.2024.

Τραυλός- Τζανετάτος Δ, Το εργατικό δίκαιο στην τέταρτη βιομηχανική επανάσταση, Ψηφιοποίηση, ρομποτική και τεχνητή νοημοσύνη, Εκδόσεις Σάκκουλα, Αθήνα – Θεσσαλονίκη, 2019, σελ. 45.

II. ΑΛΛΟΔΑΠΗ

Abernethy J., Candelon F., Evgeniou T., Gupta A., Lostonlen Y., Bring Human Values to AI, Harvard Business Review, March – April 2024, <https://hbr.org/2024/03/bring-human-values-to-ai>, τ.π. 20.02.2024.

Adams – Prassl, J., Regulating algorithms at work: Lessons for a European approach to artificial intelligence, European Labour Law Journal 2022, Vol. 13 (I) 2022, <https://journals.sagepub.com/doi/full/10.1177/20319525211062558>, τ.π. 27.03.2024.

Albaroudi E, Mansouri T, Alameer A. A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. AI. 2024; 5(1):383-404. <https://doi.org/10.3390/ai5010019>, τ.π. 24.11.2024

Aloisi A, Boss ex machina: Employer powers in workplace governed by algorithms and artificial intelligence, working paper IE Law School AJ8-274-1, 25.04.2002.

Alvarez, J.M., Colmenarejo, A.B., Elobaid, A., Policy advice and best practices on bias and fairness in AI, Ethics Inf Technol 26, 31 (2024), <https://doi.org/10.1007/s10676-024-09746-w>, τ.π. 28.08.2024.

Baesens, B., Business analytics: Analytics in a Big Data World, Wiley and SAS Business Series, 1st Edition, Wiley Editions, 2015.

- Berry, M. J. A., Linoff, G. S., Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Wiley Editions, 2004.
- Brey, P., Dainow, B., Ethics by design for artificial intelligence, AI Ethics 2023, <https://link.springer.com/article/10.1007/s43681-023-00330-4>, τ.π. 28.03.2024.
- Carter, C. (2024). Why the algorithmic recruiter discriminates: The causal challenges of data-driven discrimination. Maastricht Journal of European and Comparative Law, 0(0). <https://doi.org/10.1177/1023263X241248474>, τ.π. 24.11.2024.
- Caton S. Haas C., Fairness in Machine Learning: A Survey, ACM Computing Surveys, Volume 56, Issue 7, Article No.: 166, Pages 1 – 38, <https://doi.org/10.1145/3616865>, τ.π. 24.11.2024
- Chakraborty, J., Majumder, S., Menzies, T. Bias in machine learning software: Why? how? what to do? In Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Athens, Greece, 23–28 August 2021, pp. 429–440.
- Date S., How to Use Proxy Variables in a Regression Model., Towards Data Science, Aug 16, 2022, <https://towardsdatascience.com/how-to-use-proxy-variables-in-a-regression-model-539f723ab587>, τ.π. 31.08.2024.
- Davenport, T. H., From analytics to artificial intelligence, Journal of Business Analytics Volume 1, 2018, <https://www.tandfonline.com/doi/full/10.1080/2573234X.2018.1543535>, τ.π. 27.03.2024.
- Davenport, T. H., The new world of business analytics, Jan 2010, International Institute for Analytics.
- De Stefano V., Wouters M., AI and Digital Tools in Workplace Management and Evaluation: An Assessment of the EU's Legal Framework, European Parliamentary Research Service, PE 729516, May 2022.
- Deck L., Müller J., Braun C, Zipperling D, Kühl N., Implications of the AI Act for non-discrimination law and algorithmic Fairness, 04.01.2024, <https://arxiv.org/abs/2403.20089>, τ.π. 14.08.2024.

European Commission, Ethics By Design and Ethics of Use Approaches for Artificial Intelligence (1.0), European Commission - DG Research and Innovation, 25.11.2021, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf, τ.π. 28.03.2024.

European Parliament / Panel for the Future of Science and Technology (STOA) 2020-12 (STOA), AI and the future of work, [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656305/EPRS_STU\(2020\)656305_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656305/EPRS_STU(2020)656305_EN.pdf), τ.π. 27.03.2024.

Fundamental Rights Agency. (2022). Bias in algorithms, artificial intelligence, and discrimination Luxembourg Publications Office. <https://fra.europa.eu/en/publication/2022/bias-algorithm>, τ.π. 16.11.2024.

Güneş, F., Tharrington, R., Abbey, R., & Hunt, X. (n.d.). How to explain your black-box models in SAS® Viya®. SAS Institute Inc. <https://support.sas.com/resources/papers/proceedings20/4502-2020.pdf>, τ.π. 17.11.2024.

Hacker, P., Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law, Common Market Law Review, volume 55, Issue 4, 2018, pp. 1143-1186, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3164973, τ.π. 23.02.2024.

Hassija, V., Chamola, V., Mahapatra, A. et al. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. Cogn Comput 16, 45–74 (2024). <https://doi.org/10.1007/s12559-023-10179-8>, τ.π. 24.11.2024

Houser K., Can AI solve the diversity problem in the tech industry? Mitigating noise and bias in employment decision – making, 22 STAN. TECH. L. REV. 290 (2019).

Information Commissioner’s Office (ICO), Rights related to automated decision making including profiling, <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/rights-related-to-automated-decision-making-including-profiling/>, τ.π. 02.04.2024.

- Information Commissioner's Office (ICO), What about fairness, bias and discrimination?», <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-about-fairness-bias-and-discrimination/>, τ.π.28.03.2024.
- Jayanti L.P.S.D., Wasesa, M., Application of Predictive Analytics to improve the hiring process in a Telecommunications Company, *Jurnal CoreIT*, Vol.8, No.1, June 2022, https://www.researchgate.net/publication/362722810_Application_of_Predictive_Analytics_To_Improve_The_Hiring_Process_In_A_Telecommunications_Company, τ.π. 27.03.2024.
- Kammerer B., *Hired by a Robot: The Legal Implications of Artificial Intelligence Video Interviews and Advocating for greater protection of job applicants*, Volume 107, *Iowa Law Review* 817 (2022)
- Klimberg R., K. Miori, V., (October 2010), *Back in Business*, https://www.researchgate.net/publication/305958445_Back_in_Business, τ.π. 27.03.2024.
- Konle-Seidl, R., Danesi, S., European Parliament, Directorate-General for Internal Policies of the Union, *Digitalisation and changes in the world of work – Literature review*, European Parliament, 2022, <https://data.europa.eu/doi/10.2861/291260>, τ.π. 27.03.2024.
- Koutsoviti Koumeri L., Legast M., Yousefi Y., Vanhoof K, Legay A, and Schommer C. 2023. *Compatibility of FairnessMetrics with EU Non-Discrimination Laws: Demographic Parity & Conditional Demographic Disparity*. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.48550/arXiv.2306.08394>, τ.π. 24.11.2024
- Kuśmierczyk, M., *Algorithmic Bias in the Light of the GDPR and the Proposed AI Act (08/05/2022)*. Maciej Olejnik, Wiktoria Morawska (ed.), *"(In)equality. Faces of modern Europe"*, Wydawnictwo Centrum Studiów Niemieckich i Europejskich im. Willy'ego Brandta, Wrocław, 2022 (forthcoming), Available at SSRN: <https://ssrn.com/abstract=4117936> or <http://dx.doi.org/10.2139/ssrn.4117936>, τ.π. 02.04.2024.
- Lazcoz, G. (2021). *Automated Decision-Making Under Amsterdam's District Court Judgements: Drivers v. Uber and Ola*,

- https://www.researchgate.net/publication/356531942_Automated_Decision-Making_Under_Amsterdam's_District_Court_Judgements_Drivers_v_Uber_and_Ola, τ.π. 29.08.2024.
- Lukács A., Váradi S, GDPR-compliant AI-based automated decision-making in the world of work, *Computer Law & Security Review*, Volume 50, September 2023, 105848, <https://www.sciencedirect.com/science/article/pii/S0267364923000584>, τ.π. 27.03.2024.
- Madinier Salis F., A guide to artificial intelligence at the workplace – Your rights on algorithms, European Economic Committee, Report, 2021.
- Marín Díaz G., Galan Hernandez, J.J., Galdon, J. L., Analyzing Employee Attrition Using Explainable AI for Strategic HR Decision-Making, https://www.researchgate.net/publication/375741559_Analyzing_Employee_Attrition_Using_Explainable_AI_for_Strategic_HR_Decision-Making, τ.π. 27.03.2024.
- Noble S., U., Algorithms of Oppression: How Search Engines Reinforce Racism, New York University Press, 2018.
- O' Neil C., Weapons of math destruction, How big data increases inequality and threatens democracy, Penguin Books Ltd, 2017.
- OECD (2024), Using AI in the workplace: Opportunities, risks and policy responses», OECD Artificial Intelligence Papers, No. 11, OECD Publishing, Paris, <https://doi.org/10.1787/73d417f9-en>
- Portillo Chavez K., Bahr J., Vartanian T., 06.12.2022, AI has made its way to the workplace. So how have laws kept pace?, <https://oecd.ai/en/work/workplace-regulation-2022>, τ.π. 02.05.2024.
- Reagan M., Understanding Bias and Fairness in AI Systems, An illustrated introduction to some of the basic concepts of a crucial problem, published in Towards Data Science, 25/03/2021 <https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3>, τ.π.28.03.2024.
- Salvi del Pero, A., P. Wyckoff and A. Vourch (2022), Using Artificial Intelligence in the workplace: What are the main ethical risks?, OECD Social, Employment and Migration Working Papers, No. 273, OECD Publishing, Paris, <https://doi.org/10.1787/840a2d9f-en>.

Turing, A.M. (1950) Computing Machinery and Intelligence. Mind, 59, p. 433-460, <https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf>, τ.π. 30.01.2023.

UNESCO (2021), Recommendation on the Ethics of Artificial Intelligence», <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>, τ.π. 27.03.2024.

Xiang L., Murphy D., Applying trustworthy AI Framework to mitigate bias and increase workforce gender diversity, 2022 IEEE International Symposium on Technology and Society (ISTAS).

Zafar M., Valera I., Rogriguez M., Gummadi K, Fairness Constraints: Mechanisms for Fair Classification. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR 54:962-970, 2017.

Δελτία τύπου – Ανακοινώσεις Οργάνων ΕΕ

Ευρωπαϊκή Επιτροπή (Χ.η.), Ευρωπαϊκή προσέγγιση της τεχνητής νοημοσύνης, <https://digital-strategy.ec.europa.eu/el/policies/european-approach-artificial-intelligence>, τ.π. 27.03.2024.

Ευρωπαϊκή Επιτροπή (21.04.2021), Μια Ευρώπη έτοιμη για την ψηφιακή εποχή: η Επιτροπή προτείνει νέους κανόνες και δράσεις για την αριστεία και την εμπιστοσύνη στην τεχνητή νοημοσύνη, https://ec.europa.eu/commission/presscorner/detail/el/ip_21_1682, τ.π. 02.04.2024.

Ευρωπαϊκό Κοινοβούλιο (09.09.2020), Τι είναι η τεχνητή νοημοσύνη και πώς χρησιμοποιείται, <https://www.europarl.europa.eu/news/el/headlines/society/20200827STO85804/ti-einai-i-techniti-noimosuni-kai-pos-chrisimopoeitai>, τ.π. 25.01.2024.

Ευρωπαϊκό Κοινοβούλιο (28.09.2020), Τεχνητή νοημοσύνη: Ευκαιρίες και απειλές, <https://www.europarl.europa.eu/news/el/headlines/society/20200918STO87404/techniti-noimosuni-eukairies-kai-apeiles>, τ.π. 28.03.2024.

Ευρωπαϊκό Κοινοβούλιο (09.06.2023), Πράξη Τεχνητής Νοημοσύνης της ΕΕ: Πρώτος κανονισμός για την τεχνητή νοημοσύνη,

<https://www.europarl.europa.eu/topics/el/article/20230601STO93804/praxi-technitis-noimosunis-tis-ee-protos-kanonismos-gia-tin-techniti-noimosuni> , τ.π. 27.03.2024.

Ευρωπαϊκό Συμβούλιο (13.12.2023), Δικαιώματα των εργαζομένων σε πλατφόρμες: Συμφωνία Συμβουλίου-Κοινοβουλίου, <https://www.consilium.europa.eu/el/press/press-releases/2023/12/13/rights-for-platform-workers-council-and-parliament-strike-deal/>, τ.π. 02.04.2024.

Ευρωπαϊκό Συμβούλιο (11/03/2024), Platform workers: Council confirms agreement on new rules to improve their working conditions», <https://www.consilium.europa.eu/en/press/press-releases/2024/03/11/platform-workers-council-confirms-agreement-on-new-rules-to-improve-their-working-conditions/>, τ.π. 06.04.2024.

Ελληνικός και Ξένος Διαδικτυακός Τύπος και Αρθρογραφία

Δρακάκης Φ. (14.03.2024), AI Act: Η Ευρώπη θέλει να γίνει το παγκόσμιο πρότυπο για αξιόπιστη τεχνητή νοημοσύνη», [2045.gr](https://www.2045.gr), <https://www.2045.gr/thematologia/politiki/ai-act-i-evropi-thelei-na-ginei-to-pagkosmio-protypo-gia-axiopisti-techniti-noimosyni/>, τ.π. 02.04.2024.

Εργαζόμενοι σε πλατφόρμες: Συμφωνία για νέα Οδηγία της ΕΕ για τη βελτίωση των συνθηκών εργασίας τους (15/12/2023), Lawspot.gr, <https://www.lawspot.gr/nomika-nea/ergazomenoi-se-platformes-symfonia-gia-nea-odigia-tis-ee-gia-ti-veltiosi-ton-synthikon>, τ.π. 02.04.2024.

Κουμεντάκης Σ. (12.12.2021), Απαγόρευση διακρίσεων στην εργασία», [koumentakis.law.gr](https://koumentakislaw.gr), <https://koumentakislaw.gr/arthra/apagoefsh-diakriseon-sthn-ergasia/>, τ.π. 28.03.2024.

Οι μηχανές είναι ήδη εδώ: 4.000 άτομα απολύθηκαν μόνο τον Μάιο λόγω της τεχνητής νοημοσύνης (02/06/2023), Moneyreview.gr, <https://www.moneyreview.gr/business-and-finance/international/113422/oi-michanes-einai-idi-edo-4-000-atoma-apolythikan-mo-ton-maio-logo-tis-technitis-noimosynis/>, τ.π. 28.03.2024.

Τα εργαλεία τεχνητής νοημοσύνης μπορεί να «απορρίπτουν» τους καλύτερους υποψήφιους για θέση εργασίας (16/02/2024), [naftemporiki.gr, https://www.naftemporiki.gr/society/1592935/ta-ergaleia-technitis-noimosynis-mporei-na-aporriptoyn-toys-kalyteroys-ypopsifioys-gia-thesi-ergasias/](https://www.naftemporiki.gr/society/1592935/ta-ergaleia-technitis-noimosynis-mporei-na-aporriptoyn-toys-kalyteroys-ypopsifioys-gia-thesi-ergasias/), τ.π. 27.03.2024.

Τεχνητή Νοημοσύνη: Οι αλλαγές στη νομοθεσία - Έξι ειδικοί εξηγούν (17.03.2024), [Sofokleousin.gr, https://www.sofokleousin.gr/texniti-noimosyni-oi-allages-sti-nomothesia-eksi-eidikoi-eksigoun](https://www.sofokleousin.gr/texniti-noimosyni-oi-allages-sti-nomothesia-eksi-eidikoi-eksigoun), τ.π. 02.04.2024.

Aerts A., De Koster P., Hey T., (20.03.2024), Update: EU Directive on platform work, [twobirds.com, https://www.twobirds.com/en/insights/2024/global/update-eu-directive-on-platform-work](https://www.twobirds.com/en/insights/2024/global/update-eu-directive-on-platform-work), τ.π. 02.04.2024.

Ahramovich, A. (12.09.2024), AI in the workplace: 10 key use cases, benefits, and challenges, [intransition.com, https://www.itransition.com/ai/workplace](https://www.itransition.com/ai/workplace), τ.π. 27.03.2024.

Barry, E. (12.10.2021), Uber Drivers say a “Racist” Algorithm is putting them out of work, [time.com, https://time.com/6104844/uber-facial-recognition-racist/](https://time.com/6104844/uber-facial-recognition-racist/), τ.π. 21.02.2024.

Botelho T. L., Sudhir K., Teng F., Written by Murray S (14.08.2023), Ratings Systems Amplify Racial Bias on Gig-Economy Platforms, 14, [yale.edu, https://insights.som.yale.edu/insights/ratings-systems-amplify-racial-bias-on-gig-economy-platforms](https://insights.som.yale.edu/insights/ratings-systems-amplify-racial-bias-on-gig-economy-platforms), τ.π. 27.03.2024.

Bourgery - Gonse T. (11.04.2024), At long last, EU countries adopt the platform work directive, [euractiv.com, https://www.euractiv.com/section/economy-jobs/news/at-long-last-eu-countries-adopt-the-platform-work-directive/](https://www.euractiv.com/section/economy-jobs/news/at-long-last-eu-countries-adopt-the-platform-work-directive/), τ.π. 02.04.2024.

Bunzel C., Boon C., Algorithmic Management in Organizations: Benefits, Challenges, and Best Practices, [aihr.com, https://www.aihr.com/blog/algorithmic-management/](https://www.aihr.com/blog/algorithmic-management/), τ.π. 21.04.2024.

Chimera A (07.04.2023), How artificial intelligence can inform decision-making Artificial intelligence (AI) is a powerful decision-making tool with the potential to transform business, but challenges remain. Here’s what you need to know, [enterpriseproject.com, https://enterpriseproject.com/article/2023/4/ai-decision-making](https://enterpriseproject.com/article/2023/4/ai-decision-making), τ.π. 30.01.2023.

Corbyn Z (03/02/2024), The AI tools that might stop you getting hired, theguardian.com
<https://www.theguardian.com/technology/2024/feb/03/ai-artificial-intelligence-tools-hiring-jobs>, τ.π. 27.03.2024.

Downs J. (07/02/2022), Auchterlonie S., Proxy Problems—Solving for Discrimination in Algorithms, lexology.com, <https://www.lexology.com/library/detail.aspx?g=bf5b6949-809e-430c-8b37-997ee566f5e2>, τ.π.28.03.2024.

Dutceac Segesten A., The definition of AI, course AI, Business & the Future of Work, Lund University, Coursera.org, <https://www.coursera.org/learn/ai-business-future-of-work/lecture/MAOQw/the-definition-of-ai>, τ.π.27.03.2024.

ETUC Press release (09/12/2023), AI Act sets stage for Directive on AI systems in the workplace, etuc.org, <https://www.etuc.org/en/pressrelease/ai-act-sets-stage-directive-ai-systems-workplace> τ.π. 02.04.2024.

EU's AI Act published in Official Journal, transition periods now known (12.07.2024), cms-lawnow.com, <https://cms-lawnow.com/en/ealerts/2024/07/eu-s-ai-act-published-in-official-journal-transition-periods-now-know>, τ.π. 12.07.2024.

Gillis A., Definition of machine learning bias (AI bias), Techtargget.com, <https://www.techtargget.com/searchenterpriseai/definition/machine-learning-bias-algorithm-bias-or-ai-bias>, τ.π. 28.03.2024.

Glover, E. (29/09/2022), Strong AI vs. Weak AI: What's the difference?, Buletin.com, <https://bultin.com/artificial-intelligence/strong-ai-weak-ai>, τ.π. 27.03.2024.

Gray D. (17.04.2024), Addressing bias in AI systems through the AI Act, Mishcon.com, <https://www.mishcon.com/news/addressing-bias-in-ai-systems-through-the-ai-act#:~:text=While%20the%20focus%20of%20the,focusing%20on%20explainability%20and%20transparency>, τ.π. 02.04.2024.

HR και τεχνητή νοημοσύνη: Κορυφαίοι ειδικοί τονίζουν ότι οι άνθρωποι πρέπει να έχουν τον τελικό λόγο (26.08.2024), Fortunegreece.gr, , [HR και τεχνητή νοημοσύνη: Κορυφαίοι ειδικοί τονίζουν ότι οι άνθρωποι πρέπει να έχουν τον τελικό λόγο | Fortunegreece.com](https://www.fortunegreece.gr/hr-kei-techniti-noimosyni-koryfaioi-aidikoi-tonizoun-oti-oi-anthropoi-prenei-na-echoun-ton-teleko-logho), τ.π. 27.08.2024.

Fiegerman, S. (28.09.2016), Facebook, Google, Amazon create group to ease AI concerns, cnn.com, <https://money.cnn.com/2016/09/28/technology/partnership-on-ai/>, τ.π. 27.03.2024.

Five AI Ethics in HR: Striking Balance for a Fairer Workplace, (29.10.2023), hyscaler.com, <https://hyscaler.com/insights/ai-ethics-in-hr-workplace-balance/>, τ.π. 28.03.2024.

James V. (12.01.2018), Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech, theverge.com, <https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai>, τ.π. 07.04.2024.

Knight, W. (03/10/2017), Forget killer robots—bias is the real AI danger, technologyreview.com, <https://www.technologyreview.com/s/608986/forget-killer-robots-bias-is-the-real-ai-danger/>, τ.π. 27.03.2024.

Kourti P., (20.02.2024), ICO Issues Draft Guidance on Employment Records and the Recruitment and Selection Process, trilateralsearch.com, <https://trilateralresearch.com/uncategorized/ico-issues-draft-guidance-on-employment-records-and-the-recruitment-and-selection-process>, τ.π. 02.04.2024.

Moore P, Artificial Intelligence in the workplace: What is at stake for workers?, work in the Age of Data, bbvaopenmind.com, <https://www.bbvaopenmind.com/wp-content/uploads/2020/02/BBVA-OpenMind-book-2020-Work-in-the-Age-of-Data.pdf>, τ.π. 28.03.2024.

Nachiappan, A., (01/02/2024), I thought I was going mad': Former make-up artist says she lost her job after being marked down by AI tool, news.sky.com, <https://news.sky.com/story/amp/companies-increasingly-using-ai-for-recruiting-and-why-it-could-cost-you-your-job-13060905>, τ.π. 27.03.2024.

Needle F., What is AI bias?, (06.06.2023), hubspot.com, <https://blog.hubspot.com/marketing/ai-bias>, τ.π. 28.03.2024.

Nidumolu, K., (01.05.2023), Exploring AI Bias in Healthcare: From Gathering Data to Implementation, linkedin.com, <https://www.linkedin.com/pulse/exploring-ai-bias-healthcare-from-gathering-data-krishan-dev>, τ.π.27.03.2024.

Picchi A (17.04.2019), How tech's white male workforce feeds bias into AI, cbsnews.com , <https://www.cbsnews.com/news/ai-bias-problem-techs-white-male-workforce/>, τ.π. 28.03.2024.

President Biden issues executive order on safe, secure, and trustworthy Artificial Intelligence, Fact Sheet (30.10.2023), whitehouse.gov, <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>, τ.π. 02.04.2024.

Real-life Examples of Discriminating Artificial Intelligence, datatron.com, <https://datatron.com/real-life-examples-of-discriminating-artificial-intelligence/>, τ.π. 27.03.2024.

Shedding light on AI bias with real world examples, (16/10/2023), ibm.com, <https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/>, τ.π. 27.03.2024.

Six Types of AI Bias Everyone Should Know (12.10.2021), seldon.io <https://www.seldon.io/6-types-of-ai-bias>, τ.π.28.03.2024.

Taylor, P., (16/11/2023), Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2020, with Forecasts from 2021 to 2025, Statista.com, <https://www.statista.com/statistics/871513/worldwide-data-created/>, τ.π. 25.01.2024.

Van Bekkum M., Zuiderveen Borgesius F., (21/2/2024), The AI Act's debiasing exception to the GDPR, iapp.org, <https://iapp.org/news/a/the-ai-acts-debiasing-exception-to-the-gdpr/>, τ.π. 01.05.2024.

Wachter S, Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI, Oxford Internet Institute, University of Oxford online webinar, <https://www.oii.ox.ac.uk/news-events/events/why-fairness-cannot-be-automated-bridging-the-gap-between-eu-non-discrimination-law-and-ai/> , τ.π.16.03.2024.

Wang S, Santinelli M., Hua G., Ittner J., Mills S., (11.03.2021), Practice AI Responsibly with Proxy Variable Detection, medium.com, <https://medium.com/bcggamma/practice-ai-responsibly-with-proxy-variable-detection-42c2156ad986>, τ.π. 10.08.2024.

Yurushkin M., How Can Synthetic Data Solve the AI Bias Problem?, broutonlab.com,
<https://broutonlab.com/blog/ai-bias-solved-with-synthetic-data-generation>, τ.π.
28.03.2024.